BEYOND DYNAMIC RISK FACTORS: TOWARDS A COMPREHENSIVE
EXPLANATION OF OFFENDING

BY

ROXANNE JAYNE HEFFERNAN

A thesis

Submitted to Victoria University of Wellington

In fulfilment of the requirements for the degree of

Doctor of Philosophy in Forensic Psychology

Victoria University of Wellington

2020

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Tony Ward, for his encouragement and guidance over the past few years. It has been the greatest learning experience I could hope for and I have enjoyed every part of it, I could not have done it without your enthusiasm and belief in our project.

Next, I would like to thank members of the Explanation of Psychopathology and Crime (EPC) lab for the interesting conversations and feedback, and especially Annalisa Strauss-Hughes for sharing an office with me throughout the process and being a sounding board for all my ideas and the occasional frustration! I would also like to thank the staff and students on the forensic programme for their support and advice, in particular Louise Dixon, Clare-Ann Fortune, Daniel Wegerhoff, and Fiona Dempsey. I have learned so much from each of you and it has been a privilege to work alongside such intelligent and compassionate people – I hope we can continue this throughout our careers.

Thank you to my wonderful friends and family for the on-going support and patience throughout my time at University, it has been a long process. In particular a huge thank you to Sarah Robson for reading this entire thing and even understanding it! Last, but definitely not least, thank you to my amazing partner, Rob Willis. I couldn't ask for a more positive and caring person to be beside me throughout this journey, you truly are the best.

PUBLICATIONS INCLUDED IN TEXT

Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective factors. *Aggression and Violent Behavior (37),* 129-141. doi.org/10.1016/j.avb.2017.10.003

Heffernan, R., & Ward, T. (2019). Dynamic risk factors, protective factors, and value-laden practices. *Psychiatry, Psychology and Law, 26*(2), 312-328. doi.org/10.1080/13218719.2018.1506721

Heffernan, R., Ward, T., Vandevelde, S., Van Damme, L. (2019). Dynamic risk factors and explanation: The risk-causality method. *Aggression and Violent Behavior, 44,* 47-56. doi.org/10.1016/j.avb.2019.06.004

Heffernan, R., Wegerhoff, D., & Ward, T. (2019). Dynamic risk factors: Conceptualization, measurement, and evidence. *Aggression and Violent Behavior (48),* 6-16. doi.org/10.1016/j.avb.2019.06.004

Elsevier Permission Guidelines State: Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes. This applies to all papers published in Aggression and Violent Behavior.

Psychiatry, Psychology and Law: Permission given by Editor in Chief, Professor Ian Freckleton via email (5th September 2019).

Abstract

The current conceptualisation of dynamic risk factors (DRF) for criminal offending is problematic. There have been significant conceptual issues highlighted in this domain, however, until recently addressing these has not been a priority for researchers. Instead, research has predominantly focused on the success of DRF in predicting reoffending and the effectiveness of treatment programmes that target these factors. DRF are typically defined as aspects of individuals and their environments that are associated with an increased likelihood of reoffending, and they are widely considered 'plausible causes' of criminal behaviour. It is acknowledged that this definition encompasses a wide range of individual characteristics, social processes, behaviours, and environmental features, and that these vary in their ability to explain and predict offending. The more recent interest in features that reduce risk has prompted similar discussions about the concept of protective factors (PF). Given the frequent use of, and interest in, these foundational concepts it is timely to investigate them in depth, and to address two key issues. First, both DRF and PF are broad category labels that encompass a diverse (and largely unspecified) range of psychological and contextual features and processes. Second, without a clear understanding of what exactly these constructs are, it is difficult to effectively link them to correctional research and practice. I will begin this thesis by setting out the problems with the reliance on DRF to explain offending. I will do this by exploring recent empirical findings concerning their relationship with recidivism and outlining numerous conceptual problems which make DRF poor candidates for causal explanation. I will then suggest a shift in focus, from these crime correlates to human nature and agency, and argue that this perspective is essential in explaining any behaviour. I will present a preliminary model based on agency and demonstrate the utility of this perspective in reconceptualising DRF as aspects of goal-directed behaviour. Next, I will develop a framework for continuing this theoretical research and adding depth to theories of agency. Finally, I will discuss the implications of agency theories for forensic interventions, including their integration with widely used rehabilitation models. I will conclude with an evaluation of the approaches developed throughout this thesis and make some suggestions for future research. This research holds promise in directing the field away from the otherwise inevitable theoretical 'dead end'.

TABLE of CONTENTS

LIST of FIGURES

LIST of TABLES

Introduction

Dynamic risk factors (DRF) in their various forms are arguably the most utilised concepts in the field of forensic and correctional practice. The identification and measurement of DRF in individuals who have committed crimes is an overriding concern for the criminal justice system. They drive the formation of correctional policy and the funding of treatment programmes in prisons, forensic hospitals, and probation services. Thus, our understanding of these constructs directly influences our ability to prevent the immense social and financial costs of crime. Throughout this thesis I will argue that in their current state DRF are unable to adequately guide forensic practice, and that significant theoretical work is required in order to transform DRF into useful explanatory tools. My aim is to make some preliminary steps towards a more comprehensive understanding of offending and provide suggestions for the development of theory which utilises but does not rely solely upon DRF. The development of theory in this area will be an ongoing challenge for researchers, as will its integration within practice domains including rehabilitation programmes.

I will begin by outlining the definition, conceptualisation, and current role of DRF and forensic research and practice domains. Of primary concern for this thesis is the dual roles of DRF in practice, both as predictors of risk and as *potentially causal* propensities to be targeted in treatment. The latter role requires an explanation of a causal link between DRF and offending, as theories guide the development and delivery of treatment programmes aiming to reduce DRF. I suggest that the importation of predictors to the realm of treatment is a result of confusing correlation with causation, at least implicitly. The assumption is that because DRF statistically predict recidivism, they must be causally linked. It follows from this assumption that DRF should be targeted in treatment in order to change behaviour and should also be the focus of explanations of offending. The first chapter addresses this issue and argues that the reliance upon DRF as explanations for offending and targets for rehabilitation is misguided. This will set the scene for the development of alternative explanations which can account for the frequent co-existence of DRF and recidivism without assuming a causal relationship.

The second chapter outlines the empirical criteria put forward in the literature to determine the status of DRF as causal factors, focusing on the measurement of DRF change and recidivism (Heffernan, Wegerhoff, & Ward, 2019). The aim of this chapter is to examine the extent to which the current body of evidence supports their use in treatment, or more precisely, whether there is sufficient evidence linking DRF change with reductions in recidivism. The third chapter focuses on additional criteria for determining causality, those which are concerned with coherence and explanatory value. I will offer some conceptual

criticisms of the DRF construct, describe influential theories of offending which rely upon DRF, and then outline several recent developments in theories of offending which go beyond DRF.

The fourth chapter builds upon these criticisms; outlining the role of theory development and evaluation within scientific inquiry and the need for methodological frameworks to guide research in this field. This chapter will present an exemplar (i.e., prototypical set of DRF) which will provide examples to be drawn from throughout the rest of the thesis. While the criticisms, arguments, and suggestions apply to the correlates of illegal behaviour in general, the examples used throughout this thesis will draw mainly from the sexual offending literature and will use the pronoun "he/him" although they apply equally to individuals of all gender identities. Chapter four concludes by shifting the focus from DRF to the types of normative goal-directed practices within which they manifest, for example seeking a partner, coping with anger, and gaining resources (Heffernan & Ward, 2019). This shift begins to address some of the issues with DRF, in particular their reliance on values and norms, and their ecological sensitivity (i.e., they may be more or less adaptive in various contexts).

The fifth chapter presents a preliminary model depicting the capacities and processes underlying these goal-directed practices, the *predictive agency model* (PAM; Heffernan & Ward, 2017). The sixth chapter utilises this model within a methodological framework which aims to add depth to our understanding of DRF and offending, the *risk-causality method* (RCM; Heffernan, Ward, Vandevelde, & Van Damme, 2019). This framework represents one potential way forward; utilising DRF in explanations of offending by first breaking them down into their possible causal processes, analysing these across a range of explanatory levels, and then integrating them within the PAM. Importantly, this chapter draws together the recent literature in order to move the field forwards; it builds upon an emerging body of theoretical and meta-theoretical work rather than 'reinventing the wheel'. The seventh chapter offers some preliminary implications for forensic practice, including using the PAM and its capacities to add depth to case formulation and agency-based approaches to treatment. While theory evaluation and refinement are on-going tasks, there are some potential implications for practice at present. The eighth chapter of this thesis will conclude with an evaluation of the approaches developed throughout this thesis and some suggestions to further this body of research.

Chapter One: Dynamic Risk Factors

Meta-analyses undertaken in the 1980s and 1990s identified a core set of risk factors for criminal behaviour (Andrews & Bonta, 2010; Bonta & Andrews, 2017). According to Andrews and Bonta (2010), these risk factors are crime correlates and predictors (i.e., variables statistically associated with reoffending), including *dynamic* predictors; a subset of these are thought to be potential *causes* of crime. Proposed causal dynamic predictors have been labelled 'criminogenic needs' (Andrews & Bonta, 2010). In their book *The Psychology of Criminal Conduct* (PCC), Andrews and Bonta (2010) state that criminogenic needs are DRF which "when changed, are associated with changes in the probability of recidivism" (p. 49). Further, they specify that "dynamic predictors of criminal conduct or *criminogenic need* factors have great practical relevance because they inform interventions that reduce criminal behaviour by identifying the targets of treatment" (p. 20, italics in the original). Andrews and Bonta (2010) identified eight 'central' risk factors (seven of which are dynamic) which have the most empirical support in predicting reoffending: a history of criminal behaviour (a static factor), antisocial attitudes, antisocial associates, antisocial personality pattern, marital and family factors, employment and school, leisure activities, and substance abuse. Andrews and Bonta (2010) acknowledge that these domains can be conceptualised in different ways by different researchers, for example as psychopathic traits or weak self-control.

DRF are *changeable* features of individuals and their environments and are therefore contrasted with 'static' factors, which are risk correlates unable to be changed via intervention (e.g., criminal history, age, gender). Because of this defining feature, DRF are used as both predictors of risk and targets for change aimed at reducing instances of crime. These dual roles are reflected in the evolution of risk assessment tools from containing primarily static markers of risk to incorporating dynamic variables which can be used to guide practice. The recruitment of DRF to explain crime, formulate cases, and inform treatment is now standard practice. DRF are often further divided into 'stable' factors (i.e., enduring characteristics) and rapidly changing 'acute' factors (Hanson & Harris, 2000). For example, an individual may have a propensity to abuse alcohol or to engage in violence to solve problems (stable factors), and thus risk of further offending could be increased when he is intoxicated or experiencing conflict in his close relationships (acute factors). The stable versus acute distinction is most useful for assessment aimed at managing imminent risk and has also been conceptualised in terms of offence related vulnerability versus its manifestation in certain contexts (Beech & Ward, 2004). This view of DRF is known as a 'propensities model'; suggesting that individuals display common crime related dispositions in a variety of functional domains and that the emergence

of criminal behaviour depends on both vulnerabilities and opportunities (Mann, Hanson, & Thornton, 2010).

The most influential rehabilitation framework guiding forensic practice is Andrews and Bonta's (2010) *risk-need-responsivity* (RNR) model, which is presented in the PCC and draws heavily upon the concept of DRF or criminogenic needs. According to the RNR model, risk can be predicted, rehabilitative resources should be preferentially directed towards individuals who score highly on various risk assessment scales (risk principle), and treatment should prioritise criminogenic needs (need principle). In addition, treatment programmes should be responsive to the characteristics of the person as well as being based upon empirically supported theory and emerging evidence of the effectiveness of interventions (specific and general responsivity principles respectively). There are additional principles within the RNR model, for instance, it is appropriate to exercise professional discretion for specific reasons related to persons and their needs. In addition, 'non-criminogenic' or weakly associated factors (e.g., mental disorder) are addressed as responsivity factors in order to accommodate the complexity of human functioning. According to the RNR model, practitioners should "adapt the style and mode of service according to… relevant characteristics of individual offenders, such as their strengths, motivations, preferences, personality … and other factors" (Andrews & Bonta, 2010, p. 46). This requires practitioners to accurately assess or measure risk factors (and other features deemed relevant to offending) in order to be able to direct resources and identify needs for intervention. I will now elaborate upon the dual roles of DRF in practice, and then introduce the related concept of protective factors (PF).

1.1. Risk Assessment

Due to their strong statistical association with recidivism, the primary role of DRF in practice is to predict the probability of reoffending, and thus to determine the allocation of management and treatment resources (i.e., the risk principle). When combinations of certain features are observed to be present, rates of reoffending tend to be higher; therefore, they work as statistical predictors. The development of risk assessment tools encompasses four generations (Andrews & Bonta, 2010) evolving over the past few decades. The *first generation* utilised unstructured professional judgment, until it was found that overall these predictions were no better than chance (Hanson & Morton-Bourgon, 2009; Hart, 2009). The *second generation* saw the emergence of 'actuarial' assessment, using predominantly static risk factors to estimate the probability of reoffending in numerical form. These were more successful in predicting reoffending; however, they were severely limited in their ability to inform treatment

and the day-to-day management of individuals who have committed offences, primarily because static risk measures did not include factors that could change over time (Andrews & Bonta, 2010; Beech & Craig, 2012). The *third generation* of risk assessment measures aimed to overcome these issues by including DRF. This involves the use of various dynamic risk assessment tools as well as some subjective assessment on the part of the practitioner (Hanson & Harris, 2000; Hart & Logan, 2011). The *fourth generation* of measures extends structured professional judgment to include the use of dynamic risk assessment to inform decisions relating to case management, for example the *level of service/case management inventory* (LS/CMI; Andrews, Bonta, & Wormith, 2010). The domains covered in these tools typically coincide with the central eight risk factors (Andrews & Bonta, 2010, 2017), and in this sense they are more theoretically informed than actuarial assessment. It has been suggested that a potential *fifth generation* would use complex analyses and machine learning to predict risk in the moment (Lovins, Latessa, May, & Lux, 2018).

Numerous risk measurement scales have been constructed to aid practitioners' decision-making and ensure that risk assessment is as objective and reliable as possible (Cording & Beggs Christofferson, 2017). These measures differ by whether they focus solely on static or dynamic factors, stable and acute factors, and whether they include PF or strengths. They also vary with respect to the purposes they are used for (e.g., managing and responding to changes in risk, treatment planning, and measuring changes post-treatment), and the type of offence they are designed to predict (e.g., violent, sexual, and general recidivism). Risk prediction tools are constructed using a variety of statistical and psychometric techniques to combine factors known to correlate with reoffending. These scales are additive, and it is suggested that they include only items which increase predictive accuracy, cover multiple domains of risk, include static and dynamic predictors, and that their purpose should extend to informing the intensity and goals of treatment aimed at reducing reoffending (Cording & Beggs Christofferson, 2017). The most widely used tools all demonstrate the ability to predict reoffending at a level better than chance predictions. However, there have been concerns about their construct validity, and calls for more theoretically driven research to inform practice (Cording, Beggs Christofferson, & Grace, 2016; Kroner, Mills, & Reddon, 2005). The measurement of risk will be discussed in more detail in the following chapter, where I will outline methods and challenges in risk assessment.

1.2. Treatment Targets

As discussed above, in addition to predicting and managing risk, the second role of DRF in forensic practice is to identify treatment targets (i.e., the need principle). Correctional rehabilitative programmes are built upon the assumption that reducing or replacing DRF and strengthening or providing PF through therapy will decrease the likelihood that individuals will reoffend. While researchers often acknowledge that DRF are *potential* causes, practitioners are effectively treating them as causes when they make them the focus of treatment. Ideally rehabilitation should be preceded by careful case formulation which outlines the internal and external features which are relevant before, during, and after an offence. This process is typically based upon DRF categories, for example determining whether the individual holds attitudes which support offending, whether antisocial associates were present, and whether drugs and alcohol were involved. Forensic case formulation is based upon DRF and so these become the targets for cognitive-behavioural interventions. Indeed, programmes are often designed as a series of modules centred upon these constructs, for example 'offence-related thinking' and 'relationship issues'. Their presumed utility in case formulation and treatment rests upon their causal status.

The following two chapters of this thesis develop the argument that the causal status of DRF is currently uncertain, which means that therapists are inadvertently relying on concepts that lack explanatory power. A possible counter argument is that because interventions based on assessment of DRF result in lower recidivism rates, then the question of their theoretical coherency is moot. I think this is shortsighted. First, while correctional treatment is generally considered 'effective' in reducing recidivism, its effects vary across studies and are relatively modest overall, and so there is room for improvement (Hough, 2010). For example, Jolliffe and Farrington's (2007) international meta-analysis found an eight to 11 per cent difference in general recidivism and seven to eight per cent difference in violent recidivism between treated and untreated groups, with higher quality studies finding weaker effects. In terms of treatment for sexual offending, Schmucker and Lösel's (2015) international meta-analysis found recidivism rates of 10.1 per cent for those who completed treatment, compared to 13.7 per cent for those who did not receive treatment (representing a relative reduction of 26.3 per cent). Second, it has proven difficult to identify meaningful shifts in DRF during and following treatment (Cording et al., 2016), this is explored in chapter two. Third, as in medicine, *mechanistic explanations* of crime and reoffending are likely to result in greater awareness of possible intervention targets (Thagard, 1999). These points provide support for the scientific

and practical value of constructing richer conceptualisations of DRF and PF, and clarifying their role in explanations of offending.

1.3. Protective Factors

The field of correctional psychology has become increasingly interested in factors that *decrease* the probability of reoffending. This makes sense considering that the goal of correctional practice is to reduce the likelihood that individuals will harm others following punishment and/or treatment. Until recently, the most comprehensive investigation into PF in adults who have committed offences was a doctoral thesis by de Vries Robbé (2014), subsequently published as a series of journal articles. This body of work examined the psychometric properties of the *Structured Assessment of PROtective Factors for violence risk* tool (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009; de Vogel, de Vries Robbé, de Ruiter, & Bouman, 2011), while also discussing what PF are, and proposing how they might exert their positive effects. De Vries Robbé, Mann, Maruna, and Thornton (2015) define a PF as "a feature of a person that lowers the risk of reoffending" (p. 18) and propose that "the definition of a protective factor should encompass social, interpersonal, and environmental factors as well as psychological and behavioural features" (p. 18). In line with the conceptualisation of DRF, they differentiate between a PF as an underlying propensity (i.e., a psychological characteristic) versus an observable manifestation of that propensity (e.g., employment may be a manifestation of underlying propensities such as conscientiousness and social skills).

In the more recent versions of the PCC, Andrews and Bonta (2010; Bonta & Andrews, 2017) have adopted the term 'strengths' to cover the categories of protective and promotive factors discussed in the wider criminal justice literature. In brief, PF are associated with a decreased risk of offending while 'promotive factors' are associated with positive outcomes in general, regardless of the presence of risk (e.g., healthy brain development). While researchers have identified factors that protect persons from the *onset* of offending (i.e., promotive factors such as having above average intelligence and close relationships with at least one parent; Lösel & Farrington, 2012), in the forensic domain there is greater interest in factors associated with *desistance* from offending once an individual has already had contact with the criminal justice system. A range of 'desistance factors' have been discovered by researchers, for example, the concept of 'turning points' or experiences that can redirect someone towards or away from crime (Sampson & Laub, 2005). The processes and events statistically associated with desistance include marriage, changes in self-narratives, stable employment, joining the

military, and parenthood (Sampson & Laub, 2005). These factors mirror several of Andrews and Bonta's (2010) DRF, and as such may be thought of as descriptions of areas of strength or indicators of a lifestyle less compatible with crime.

This increasing attention has brought with it debate concerning what PF *actually are*, for example whether they are merely the opposite or absence of DRF (bipolar) or something different entirely (unipolar), and how they might function to reduce risk (Fortune & Ward, 2017). For example, one may suggest that these two concepts exist at opposite ends of a continuum, and therefore there is no need to distinguish between them (i.e., there is no need to consider prosocial attitudes as they are subsumed under antisocial attitudes). However, the concept of protection is important for two simple reasons (Rutter, 1987). First, even if the two types of factors exist on the same spectrum, we still need different terms to describe each pole. Arguably the concept of protection adds epistemic value in addition to the concept of risk. By distinguishing between the two poles we can determine the current direction of individuals' life trajectories as shifting towards or away from maladaptive outcomes (i.e., offending). Second, the factors involved in protection *can* differ from those in risk processes, with Rutter (1987) providing the example that being shy or introverted may protect children from delinquency but being outgoing does not predispose them to delinquent behaviour. In addition, there is debate concerning the extent to which DRF and desistance processes "belong together" (Polaschek, 2016), and some empirical studies have demonstrated differences between crime causing and reducing factors (e.g., Kroner, Polaschek, Serin, & Skeem, 2017). De Vries Robbé et al. (2015) argue that "protective factors must exist as definable propensities or manifestations thereof in their own right, rather than being no more than the absence of a risk factor" (p. 41). As above, they also divide PF into bipolar and unipolar, they term the latter 'unique' PF, for example, medication and intelligence.

In order to effectively use PF within risk assessment and intervention, researchers need to determine how PF might function to reduce risk. In other words, "instead of searching for broadly based protective factors, we need to focus on protective mechanisms and processes" (Rutter, 1987, p. 137). For example, rather than appealing to 'concern for others' as protecting individuals from engaging in crime, we should examine *how* and *why* some people develop (or fail to develop) concern for others in specific contexts. PF are hypothesised to interact with DRF in ways that modify their influence; that is, they buffer their effects (Farrington, 2016). Some factors have been described as promotive in low risk contexts, but take on more protective functions (i.e., moderate or buffer risk factors) in the face of adversity. For example,

good parenting is a promotive factor, but it may also take on a further protective function for a child experiencing adversity.

Similarly, de Vries Robbé (2014) suggests four 'mechanisms' or routes by which PF have an impact on risk. The *risk reducing effect* influences DRF directly, with the example of medication lessening the severity of symptoms of mental illness[1]. The *moderator or buffering effect* involves lessening the strength of the relationship between a DRF and offending rather than changing the DRF directly. For example, the impact of self-control on the relationship between substance use and offending. The *main effect* is more general, offering overall protection rather than influencing specific DRF. Examples include work, leisure activities, and life goals. The final mechanism is the *motivator effect*, in which PF have a positive influence on each other. For example, intelligence and secure attachment may facilitate later development of other PF, such as empathy, work, social network, and motivation for treatment. De Vries Robbé (2014) suggests that these routes are not mutually exclusive, and that PF vary by individuals, specific contexts, and over time. For example, a prosocial and supportive social network may directly reduce stress, while also acting as a moderator by supporting the individual to abstain from substances that might increase risk. Additionally, social networks may globally improve one's life, directly reducing risk and facilitating the development of other PF. This is in line with Thornton, Kelley, and Nelligen's (2017) proposal that internal strengths may be displayed within certain (prosocial or supportive) arenas throughout the desistance process. However, without a clearer understanding of how these processes occur, we are limited in our ability to influence offending related behaviour.

1.4. Summary: Dynamic Risk and Protective Factors

To summarise, DRF are correlates of offending which can (at least in theory) be altered through treatment; for this reason, they are used in both risk assessment (i.e., risk management and classification) and treatment (i.e., goals and programme content). I argue that while they may be useful for the first task, they have been prematurely imported into the domains of treatment and explanation. They require additional theoretical work in order to be useful explanatory tools, the reasons for this will be explored further in the third chapter. I have also introduced the concept of PF and it is important to spell out what this means moving forward. I will not consistently discuss PF alongside DRF, although they will be discussed at various points throughout this thesis. The reason for this is that I tentatively conceptualise DRF and PF

---

[1] It is acknowledged that mental illness is not a predictor of general recidivism; this section draws from examples offered by the authors of the SAPROF tool which was developed within a forensic mental health context.

as different aspects of the same domains of functioning and contexts, they are "families of related concepts" (Thornton et al., 2017, p. 30). It seems to me that DRF and PF are for the most part different aspects of the same core abilities and arenas, and that even those PF that on the surface appear unipolar will exist upon a continuum that corresponds to an established domain of risk. Or alternatively, that the opposite end of established DRF/PF continua have not been the focus of research in the forensic field. In addition, due to their dual nature (Fortune & Ward, 2017) these core abilities and arenas are not exclusively risky or protective, they function in different ways for various persons and contexts (e.g., social skills can either be used for seeking a consenting relationship with an adult or grooming a child). Consequently, this thesis focusses on DRF, but the same arguments and conclusions apply to PF. The rationale for this will become clear in the following chapters; significant conceptual problems leave both concepts equally flawed when it comes to explaining the causes of behaviour.

Chapter Two: Empirical Evidence

It has recently been suggested that "the theoretical legitimacy of incorporating dynamic risk factors into the domain of treatment depends on their causal status" (Ward & Fortune, 2016a, p. 80). Researchers who are interested in the relationship between dynamic risk factor (DRF) change, treatment, and recidivism are increasingly noting that the current conceptual understanding of DRF is impacting on the quality of research (Beech, Wakeling, Szumski, & Freemantle, 2016; Cording et al., 2016; Klepfisz, Daffern, & Day, 2016; Serin, Lloyd, Helmus, Derkzen, & Luong, 2013; van den Berg, Smid, Schepers, Wever, van Beek, Janssen, & Gijs, 2018). There is concern that DRF may just be 'symptoms' or 'proxies' of underlying causal processes, rather than established causes themselves (Klepfisz et al., 2016; van den Berg et al, 2018; Ward & Fortune, 2016b). More generally, it is claimed that the current conceptualisation of DRF is overly simplistic, making it difficult to understand the specific mechanisms underpinning important phenomena, such as reoffending or the change process (Serin et al., 2013). Such issues have largely arisen from the importation of predictive constructs into the realm of explanation (Ward, 2016), and it is therefore necessary to bridge the conceptual gap between empirically derived predictors and explanations of offending to better inform management and intervention (Klepfisz et al., 2016). In other words, researchers must determine whether the importation of predictive constructs into the domain of explanation is warranted. This rests upon whether they can possibly be considered causal (or at least proxies for causal processes), and therefore whether they are meaningful targets for change.

This chapter will begin to address the presumed causal nature of these constructs by reviewing the recent empirical evidence concerning DRF and change, relying upon the most comprehensive reviews and meta-analyses (see Heffernan, Wegerhoff, & Ward, 2019). The studies reviewed here are illustrative of the literature in this area and are not exhaustive. Three key questions I will explore are: what is the evidence that DRF change? Is DRF change linked with recidivism? Does treatment produce these changes in DRF? It is acknowledged that these are not the only requirements necessary to determine causality, there are also theoretical concerns (see Bradford-Hill, 1965). These will be addressed in the following chapter, after I have summarised the empirical findings. I will first outline the causal criteria put forward in the forensic literature, and the extent to which the empirical research addresses the nature of DRF. Next, I will consider the measurement of DRF and some challenges faced by researchers in this field. I will then explore each of the empirical criteria that DRF need to meet in order to be considered 'causal covariates' (i.e., the questions above), or features that tap into the causes

of crime. This evidence is crucial in supporting the need principle of the RNR model (Bonta & Andrews, 2017).

2.1. Determining Causality: Empirical Criteria

In addition to Bonta and Andrews' criteria: "(1) deliberate interventions produce changes on the potential need factor, and (2) deliberate interventions produce changes in criminal conduct" (2017, p. 24), various authors have provided suggestions for determining the causal status of risk factors (e.g., Douglas & Skeem, 2005; Heffernan, Ward, Vandevelde, & Van Damme, 2019; Kraemer et al., 1997; Mann, Hanson, & Thornton, 2010). For example, Kraemer et al. (1997) propose that a causal risk factor must satisfy three empirical requirements, 1) precede and increase risk, 2) demonstrate change (via intervention or other mechanisms of change), and 3) predict changes in behaviour when altered. Arguably DRF have met the first criteria (Bonta & Andrews, 2017), but the extent to which they meet the second and third is still under investigation and will be explored further below. In addition, Mann et al. (2010) propose that for a DRF to be 'psychologically meaningful' it must be reliably associated (i.e., statistically predict) with recidivism and be *plausibly* causal. These authors present a list of DRF for sexual offending which tentatively meet these criteria (i.e., the strongest predictors), although they acknowledge that theoretical work is required to causally link these with offending. This recognition of the importance of theory supports the idea that the criteria for determining causality are not only empirical. In addition to displaying the expected statistical relationships with offending there must be explanatory links between variables. For example, theories depicting the complex interrelationships between various causal processes and behaviour.

While causal criteria are outlined in the literature, empirical studies which rely on the operationalisation and measurement of DRF often fail to explicitly address their nature and their presumed relationship to offending. The conceptualisation of DRF is regularly overlooked in favour of some brief reference to criminogenic needs, the central eight, or the specific risk assessment tools used. This neglect is problematic, as these studies rely on DRF to guide intervention and measure treatment progress, rather than focusing solely on the prediction of reoffending. It is as if their causal status is taken for granted and so there is no need to discuss this in the rationale and design of these studies. There are exceptions, however, and empirical studies are increasingly paying some attention to conceptual issues (e.g., Mann et al., 2010; Serin et al., 2013; van den Berg et al., 2018).

2.2. The Measurement of Dynamic Risk Factors

As discussed in the previous chapter, one of the key tasks of correctional practice is the measurement of risk factors and estimation of the likelihood of recidivism via risk assessment. This can be more or less structured and objective, depending on the extent to which it relies on empirically derived risk factors or practitioner intuition. Methods include clinical interviews and professional judgment, using structured risk tools (alongside interviews) or psychometrics containing empirically derived (actuarial) predictors, using mechanical methods to produce numerical estimates, and/or a combination of these methods (see Beech et al., 2016; Helmus, 2018). The aim of structured risk assessment is to provide a systematic and unbiased method of identifying: a) an individual's risk of recidivism and by extension need for intervention, and b) the appropriate targets of intervention to reduce the likelihood of recidivism (Bonta & Andrews, 2017; Latessa & Lovins, 2010). Thus, risk assessment allows practitioners and policy makers to supervise individuals and distribute resources in the most efficient and cost-effective manner whilst also promoting community safety (Bonta & Andrews, 2017; Latessa & Lovins, 2010). To ensure the accuracy of assessments, correctional institutions generally utilise actuarial risk assessment tools which are designed to identify the presence and severity of factors statistically associated with risk (i.e., the central eight and related predictors).

2.2.1. Risk assessment tools.

Two sets of tools frequently utilised to measure DRF and their change over time are the *level of service inventory* (LSI) measures (Andrews & Bonta, 1995; Andrews, Bonta, & Wormith, 2004) and the *violence risk scale* instruments (VRS; Wong & Gordon, 1999-2003, 2006). The LSI tools were developed to measure risk of *general* reoffending and identify relevant targets for intervention, and they are based on the central eight criminogenic needs (Andrews, Bonta, & Wormith, 2004). The VRS was specifically designed to measure risk of *violent* reoffending, identify relevant treatment needs, assess 'readiness for change' across needs, and measure change during treatment (Wong & Gordon, 1999-2003; 2006). The VRS tools contain static and dynamic factors linked to violent (or sexual in the sexual offending version) reoffending, which overlap with the central eight. While these tools are most often used in the research cited in this chapter, it is acknowledged that there is variation in items across tools used for various populations and purposes. For example, risk factors for general recidivism, violence, and sexual offending overlap but there are also unique factors for each type of behaviour. Therefore, while I will discuss DRF throughout this chapter, I am essentially

referring to actuarial predictors of any type of recidivism that are theoretically capable of change, not just the central eight.

While questions have been raised about the overall predictive validity of certain tools and their individual items (e.g., the LSI; Baird, 2009; Duwe & Rocque, 2016; Giguère & Lussier, 2016), the LSI and VRS tools "have generally high levels of predictive accuracy that are at least on par with static measures of risk" (Cording et al., 2016, p. 84). One possible reason that critics and supporters of DRF measurement disagree concerning the evidence is due to their differing standards for what constitutes acceptable predictive validity (Casey, 2016). For instance, there is considerable disagreement among researchers regarding interpretations of *area under the curve* (AUC) scores and what constitutes acceptable predictive validity. With no consensus, researchers draw upon different sources to interpret AUC scores. While some researchers utilise interpretations drawn from Cohen's (1969; 1988; 1992) works (see Hanson, 2009), others opt for values which align more with academic grading or medical standards (see Baird et al., 2013; Duwe & Rocque, 2016; Thornton & Laws, 2009). In addition, it is necessary to consider the source of the evidence in question, as considerable variation has been found between evaluations conducted by tool developers and non-developers, with developers typically finding stronger support for their own risk assessment tools (i.e., the 'allegiance effect'; Andrews et al., 2011; Duwe & Roque, 2016).

The LSI, VRS, and other widely used risk assessment tools rely on empirically derived predictors, require a practitioner to assign a numerical rating to each, and then provide an overall risk score based on the sum of all factors. These scores often correspond to (largely arbitrary) risk bands, for example high, medium, or low risk. However, this has recently been advanced with the development of a common risk language containing five levels of risk and need which can guide risk management and intervention (Hanson et al., 2017). Numerical scores based on actuarial (i.e., empirically derived) items can give the recipients of risk assessment outcomes a sense that these are objective and reliable measures, absent from bias because they are based in science. However, ratings are primarily based upon professional judgments and the quality of the information available to the rater (i.e., interview, file review, observation and self-report; see Beech et al., 2016). The subjectivity of these perspectives leaves room for bias, and as such, research which relies upon risk assessment ratings should be viewed with this in mind. For example, some studies have found poor levels of agreement when scoring risk assessment tool items or inadequate testing of inter-rater reliability (Baird, 2009; Duwe & Rocque, 2016), posing a challenge to any research which relies on the measurement of DRF with such tools.

2.2.2. Measurement of DRF change.

The measurement of DRF change is crucial in determining their dynamic and causal nature. However, according to Grady, Brodersen, and Abramson (2011) in many studies:

> … changes in these targeted areas are not evaluated to determine outcome success in treatment, despite the fact that these ... are considered influential in the overall change process. As such, a wide gap exists between the distal outcome (recidivism) and the proximal outcomes. (p. 228)

In other words, research often relies on comparisons between recidivism outcomes for treated and untreated groups and concludes that reductions in recidivism for those treated can be attributed to changes in DRF caused by the intervention. However, in order to adequately support this claim, researchers must demonstrate that this relationship is mediated by changes in DRF, where these have been targeted in treatment (Beggs, 2010). This requires pre- and post-testing in order to rule out the possibility of extraneous variables (e.g., maturation and other naturally occurring events) influencing outcomes (Beggs, 2010; Bonta & Andrews, 2017; Jung & Guyalets, 2011). For example, using risk assessment tools or psychometric measures designed to tap into the targets of change (i.e., DRF such as attitudes and self-control).

There are various ways to quantify the magnitude of change during treatment. As discussed by Labrecque, Smith, Lovins, and Latessa (2014), these methods include: the use of *raw difference scores* (e.g., Vose et al., 2009) and difference scores which control for pre-treatment score (e.g., Beggs & Grace, 2011; Olver, Nicholaichuk, Kingston, & Wong, 2014), dividing a sample into *typologies* based on their initial score and the direction and strength of change between time points (a method which is particularly common in studies using the VRS, e.g., Olver & Wong, 2011), and the calculation of *percentage change* in score (e.g., Vose, Lowenkamp, Smith, & Cullen, 2009). In additional, some studies measure *clinically significant change* (e.g., Barnett, Wakeling, Mandeville-Norden, & Rakestrow, 2013; Olver, Beggs Christofferson, & Wong, 2015). Ideally this research would take the form of *randomised controlled trials* (RCTs) in order to reduce the influence of extraneous influences. However, due to practical and ethical challenges, such as selecting valid constructs/measures and withholding treatment from control groups (Jung & Guyalets, 2011; Marshall & Marshall, 2007), this research tends to be quasi-experimental. For instance, using existing treated and untreated groups and relying on matching techniques to reduce the likelihood of significant

(risk related) differences between groups. Therefore, despite the wealth of programmes centred upon DRF, quality research is limited (Beggs, 2010; Olver & Wong, 2013).

2.3. Challenges in Measuring Dynamic Risk Factors and Change

The practice of identifying and detecting changes in DRF has been the subject of debate, based on both empirical and conceptual issues. I have briefly discussed some of the methods for measuring DRF; I will now outline a number of conceptual and empirical issues which arise in this research. These issues are particularly important when it comes to considering the third question, of whether changes to DRF through treatment lead to reduced rates of reoffending.

2.3.1. Conceptual issues.

As discussed by Helmus (2018), DRF measurement tool items are assumed to be *indicators* of underlying risk-related propensities. For example, behaviour such as 'non-compliance' (e.g., failing to report for probation meetings, refusal to engage in treatment) is considered evidence of crime-supportive cognition. Helmus (2018) goes on to suggest that the same construct can be measured in different ways, for example, observing past behaviour (i.e., offending history, a static factor), identifying current propensities (i.e., personal vulnerabilities or DRF), and looking at the presence or absence of strengths in a particular domain (e.g., social supports). In other words, static, dynamic, and protective items may just be tapping into different aspects of the same constructs. The idea that static and dynamic factors are related (Casey, 2016) is in line with a propensities model of risk; for example, the notion that static factors are the result of vulnerabilities manifesting in past contexts (Beech & Ward, 2004). Ward and Beech (2015) highlight issues with the measurement and conceptualisation of DRF due to their status as 'hybrid concepts' which may not reflect unique constructs as previously thought and may instead be tapping into the same underlying psychological propensities as static risk tools. This issue is also discussed in a recent meta-analysis; van den Berg et al. (2018) found that "the incremental validity of dynamic over static risk assessment was established for all outcome measures. However, effect sizes tended to be small, which suggests that static and dynamic instruments overlap, at least when it comes to their predictive value" (p. 187). This is a popular idea, with numerous empirical and theoretical papers discussing the intersectionality of static and dynamic factors (Beech & Ward 2004; Casey, 2016; Helmus, 2018; Ward & Beech, 2015).

Another important issue for the measurement and conceptualisation of DRF is the fact that they are strongly *interrelated*; general dynamic risk domains overlap and aspects of them may interact. For example, attitudes and personality features are directly implicated in

interpersonal relationships and lifestyle factors such as employment and marital factors. While this overlap is desirable for scales designed to measure a construct (e.g., depression), Baird (2009) states that "for risk assessment, it is best when all risk items are totally independent of each other but each has a relatively strong relationship to the outcome measure utilized" (p. 8). This causes problems for risk prediction as well as explanation; statistical interrelation likely stems from poor conceptualisation of these variables. Cording et al. (2016) have similarly highlighted issues such as 'fundamental' construct validity (i.e., measuring what it is designed to measure). They suggest that the composite nature of DRF is a problem, because it is unclear what exactly is being measured and which aspects may be linked with recidivism. They further suggest that DRF should demonstrate a causal or explanatory link with recidivism in order to be considered valid constructs (Cording et al., 2016).

DRF were derived from large meta-analyses, which, by nature, aggregate group-level data to observe patterns and variation in outcomes such as recidivism (Andrews & Bonta, 2010). However, this does not align with how DRF are frequently treated in practice, as changeable features of *individuals* and/or their environments. Just because a trend is observable between individuals, does not mean it is necessarily observable *within* an individual. There are several issues that arise from attempting to justify the use of factors or tools generated from aggregate data to guide the management and treatment of individuals. It has been argued that such an approach is "conceptually incomplete and misses important nuances" (Hannah-Moffat, 2009, p. 213). As such, more recent work has recognised the importance of conducting intra-individual measurements of change to identify changeable features within a person and/or their environment (Beggs, 2010; Kroner, & Yessine, 2013; Serin et al., 2013). Further, when using DRF with individuals, it is important to determine whether or not the DRF is *relevant* to this specific person and behaviour (Baird, 2009), and this involves functionally linking the DRF with the behaviour and understanding the type and degree of association (i.e., as driving behaviour, destabilising, or disinhibiting; Klepfisz, et al., 2016). While the shift from group-level to individual-level measurement is intuitively a desirable development, problems arise due to insufficient attention to the conceptualisation of DRF.

2.3.2. Empirical issues.

In addition to these conceptual issues, researchers have critiqued the various methods used to measure DRF change and highlighted methodological discrepancies between studies. There does not yet seem to be a consensus on how best to measure changes in DRF across time, and this often leads to diverse findings. For example, Beech et al. (2016) observe that "various

methods may produce differential results in terms of both the identification of need and the amount of change observed, which complicates any attempts to assimilate evidence on dynamic risk factors" (p. 78). Relatedly, studies utilise various risk assessment tools, and so a lot of the time they are measuring different things. For example, while the LSI tools converge around the central eight, the VRS measures specific factors relating to violence and other tools measure factors predictive of sexual recidivism. More broadly, Sullivan (2019) has recently highlighted a lack of consensus regarding the conceptualisation and measurement of key constructs in the fields of criminology and forensic psychology, which she argues has severely limited our ability to develop causal explanations for phenomena of interest. Likewise, such arguments apply here. Many researchers share the common goal of investigating the relationship between DRF and outcomes such as recidivism. Yet, due to the discrepancies in how researchers define, conceptualise, and measure such constructs, we cannot be certain that they are referring to the same *things*. Subsequently, we cannot reliably generalise or compare findings across studies, nor assume that these variables are 'real things' that exist in the way we expect they do.

The literature concerning DRF change is also confusing due to the use of psychometric scales which are assumed to measure constructs related to domains of dynamic risk (e.g., attitudes, empathy), but are not necessarily equivalent to those psychometric or risk tools utilised in other DRF change studies. As Beggs (2010) describes, these studies use "psychometric change scores, risk instruments, Goal Attainment Scaling, and various other idiosyncratic rating systems" (p. 369) measuring "variables that had previously been empirically or theoretically linked to risk" (p. 374) to operationalise within-treatment change. This variation makes it difficult to infer whether a finding of no change according to risk measures or psychometric tools reflects an *actual* lack of change or treatment failure, that the measurement tools selected are not detecting change that has occurred, or that the variables chosen (DRF) are not able to change (Cording et al., 2016). In addition, the specific items used in these tools may or may not themselves be considered DRF but may instead be viewed as indicators (or proxies) of problems within a broader DRF domain (e.g., mental health history being used to identify emotional or personality problems). Therefore, each item may be more or less predictive of recidivism; this is assumed to be due to the conflation of risk and need and can cause problems for classification of risk level and allocation to interventions (Baird, 2009).

Similarly, the reliance on imprecise and often poorly operationalised variables such as 'treatment' (which can involve various tasks) and 'recidivism' (which is difficult to reliably detect) impairs our ability to draw any strong conclusions about their interrelationships. For example, in addition to issues of detectability, recidivism is often reduced to a binary 'yes/no'

outcome. It provides little insight into other offence (or desistance) related outcomes (e.g., reduced frequency or severity of offending, similar or undetected behaviours) which may tell researchers whether positive changes were made during treatment (Olver, Lewis, & Wong, 2013). In regard to treatment, while studies might compare the use of *cognitive-behavioural therapy* (CBT) approaches to other modalities, we cannot assume that all programmes based on the same theory are equivalent. For example, CBT is based upon a cognitive-behavioural model and outlines key tasks such as challenging unhelpful (i.e., antisocial in this context) thinking and practicing skills to cope with emotions, deal with conflict, and problem-solve. However, there are a number of ways in which CBT programmes can vary; from the level of training and interpersonal style of therapists, to the way the programme is structured and the focus of individual sessions, and the specific discussions and questions encountered by participants. In addition, programmes might be more or less well matched to the target population, a programme that works well in one context might not be appropriate in another. Perhaps programme heterogeneity is implicated in the diverse findings of treatment evaluation studies based on CBT programmes, some of which have found strong support (e.g., Olver & Wong, 2011) and others which have found no effect or effects in the wrong direction (e.g., Olver, Kingston, Nicholaichuk & Wong, 2014).

Further, while it is useful to know whether DRF actually do change as a result of treatment or other influences, it is also useful to know if these changes continue to occur (i.e., across more than two time points) and if they translate to differences in future behaviour. There are limitations to what researchers can infer from measuring risk at two time points only, and "a more nuanced understanding of offender change needs to disentangle true change from measurement error" (Helmus, 2018, p. 46). Thus, the measurement of change across more than two time points is important to track the magnitude of difference in scores and detect patterns in DRF change throughout interventions; as such, it may help researchers to understand how or why change is occurring (Hendry, 2013; Klepfisz et al., 2016). It is also a more robust method, as the assessment of risk at multiple time points reduces the likelihood that changes are a result of measurement error. In addition, it has been suggested that controlling for baseline risk, and thus focussing on the magnitude of change rather than risk scores themselves, is important (Beggs, 2010).

Finally, in a recent paper Yang, Guo, Olver, Polaschek, and Wong (2017) provide a critique of how statistical analyses are used to investigate change in risk. Key threats to accuracy commonly included "missing data, irregular intervals in repeat risk assessments, and individual differences such as age and risk levels" (p. 59), as well as the use of unsuitable or

'one size fits all' analyses. In order to improve future research, these authors present methodologies drawn from the fields of biostatistics and epidemiology in attempts to overcome these threats (Yang et al., 2017). While recent research has begun to explore more sophisticated statistical techniques for measuring DRF change (Helmus, 2018), researchers have only just begun to use these techniques to attempt to link DRF change with recidivism (e.g., Olver et al., 2018).

2.3.3. Implementation and biases

In addition to the above methodological challenges, there are issues with the reliability of scoring risk assessment tools. Helmus (2018) states that "it may not be enough to implement an evidence-based scale – you need to demonstrate that you have implemented it *well*" (p. 46; emphasis in the original). Cording et al. (2016) provide a thorough evaluation of change measurement methods, and discuss issues such as a:

> … lack of specialised, standardised training; high levels of work-related stress … large workloads and time pressures leading to a greater reliance on clinical or professional judgment rather than a strict adherence to scoring guidelines… [and] fear of political and professional implications of having rated someone as low risk who later goes on to reoffend (even if the rating was correct). (p. 91)

They suggest that these issues may encourage practitioners to adjust risk assessment scores in order to over-estimate risk. Beech et al. (2016) similarly highlight the subjectivity of risk ratings and a tendency to under-estimate change or over-estimate risk. In addition to the ethical problems arising from over-estimation of risk, these potential errors in ratings have implications for the measurement of DRF and change. Researchers cannot be confident that risk or the magnitude of change has not been over- or under-estimated at any time point.

Another concern for psychometric measurement or interviews which rely on self-report is a social desirability bias or 'impression management' (Beggs, 2010; Tan & Grace, 2008; Tierney & McCabe, 2001). In order to combat this issue and provide a more accurate estimation of treatment change, it has been suggested that practitioners and researchers should rely upon a triangulation of methods rather than one source alone, for example "by using objective and subjective measures of change, including behavioural indicators" (Klepfisz, O'Brien, & Daffern, 2014, p. 359), or providing some measure to detect social desirability bias. Unfortunately, this is not yet routine practice, with a review by Serin et al. (2013) finding that

90.1 per cent of the studies examined did not include any measures of social desirability or impression management. This is particularly problematic, as research has found that promising pre-post treatment changes have been heavily weakened after controlling for social desirability (e.g., Olver & Wong, 2013), suggesting that there may be a tendency for participants to present themselves more favourably following treatment (i.e., to gain positive progress reports or early parole).

2.3.4. Summary: Challenges.

It has been suggested that "further evidence is required in order to confirm how best to measure dynamic risk factors, and to confirm the notion that changes on dynamic risk factors are associated with meaningful reductions in offenders' recidivism risk" (Beech et al., 2016, p. 78). The challenges discussed here are significant and do limit the conclusions we can currently make concerning the relationship between DRF, treatment change, and recidivism. However, until we have a more co-ordinated body of research using consistent and objective methods, it is the best empirical evidence available concerning the nature of DRF and their relationship with offending behaviour.

DRF remain central to forensic practice due to the widespread assumption that they are causally linked with offending. In order to justify their continued use in their current form, the following must be established of DRF: 1) dynamic validity or changeability; 2) a consistent relationship between DRF change and meaningful outcomes such as recidivism, and; 3) treatment must be the moderator of this change. If these empirical requirements cannot be met, then perhaps DRF in their current form have limited value for intervention. If they can be met, then further theoretical work is required to elaborate on their proposed causal relationship with offending, and to better understand the change process (i.e., identifying the mechanisms underpinning offending and DRF change).

2.4. Dynamic Validity

Recognising the importance of being able to assess and respond to increases or decreases in risk, the dynamic validity of DRF is of interest to researchers and practitioners alike. Upon reviewing the literature, it is apparent that this research largely takes the form of measuring group DRF scores at two time points and conducting statistical analyses to identify whether significant changes have occurred between the two, as described in the previous section. It is worth noting that there could be different reasons for why they change, for example, maturation and age-related changes, changes to social roles and contexts, and/or the effects of treatment (Cording et al., 2016; Helmus, 2018). Dynamic validity tells us whether

these supposedly dynamic factors do in fact tend to change across time, or more accurately, whether risk assessment tools or psychometric measures detect changes. It does not tell us whether treatment influenced these changes, or if they result in different behaviours (i.e., increased or decreased rates of recidivism). It is important to note that while at face value DRF are theoretically changeable (e.g., it is clear someone who was previously unemployed can later become employed), it is necessary to ensure that tools are detecting this change, and to determine whether some factors are more changeable than others, or alternatively some may take longer to change (i.e., stable versus acute factors).

2.4.1. Evidence for dynamic validity.

Numerous studies have found that DRF scores can change across time, often these studies have investigated pre- and post-treatment scores on risk assessment tools, and so they attribute any change detected to treatment progress (e.g., Beggs, 2010; Kingston & Olver, 2018; O'Brien & Daffern; 2017; Olver & Wong, 2011; Olver, Nicholaichuk, & Wong, 2014). Additional studies have found support for the changeability of some DRF and not others (Brooks Holliday, Heilbrun, & Fretz, 2012; Greiner, Law, & Brown, 2015; Labreque et al., 2014; Schlager & Pachego, 2011; Wooditch, Tang, & Taxman, 2014), or have identified only negligible levels of change (Hildebrand & de Ruiter, 2016). However, while this may be a comment on the dynamic validity of these factors, it may also be due to a range of other variables, for example measurement error or ineffective treatment (Cording et al., 2016). Hanson, Harris, Scott, and Helmus (2007) looked at change independent of treatment and found that "offenders changed little on the stable factors during the 6 month retest period" (p. 25), suggesting that DRF may be less likely to change without intervention, at least in the short term. Conversely, Wooditch et al. (2014) concluded that their findings provided "evidence to suggest that factors other than treatment are driving need changes" (p. 284). Thus, evidence for dynamic validity varies both in the degree of change observed across different DRF and also the presumed cause of any changes identified.

2.4.2. Summary: Dynamic validity

Overall, the evidence tends to support the assumption that DRF can change, despite the existence of some studies demonstrating that only some DRF scores changed; or that some individuals demonstrated changes while others did not. Conservatively however, it is important to consider that due to the poorly defined nature of DRF and the various methods, instruments, and items used to measure them, it is unclear exactly *what* is changing and if any specific

components within these particular factors are driving this change. Such limitations therefore make it difficult to confirm the dynamic validity of these variables with certainty.

2.5. Change and Recidivism

As above, the evidence suggests that DRF scores do change and there is some support for their dynamic validity. In line with Kraemer et al.'s (1997) criteria, it is therefore necessary to demonstrate that DRF change is meaningfully linked to recidivism. Yet, according to Helmus (2018) "there has been surprisingly little research to examine whether lower scores upon reassessment are associated with reduced recidivism" (p. 46). It is worth reiterating at this point that linking DRF change with reduced recidivism is not without methodological limitations, some of which were discussed in the previous section. The issue most relevant to this section is the use of recidivism as a dependent variable. I will therefore elaborate on this slightly here.

While the use of recidivism as an outcome measure is standard practice, the way that recidivism is operationalised is inconsistent across studies and does not necessarily translate to meaningful change (Grady, Edwards, & Pettus-Davis, 2017). For example, recidivism measures often rely on official records of rearrests, reconviction, or re-incarceration; however, some studies include self-report measures. This means that instances of recidivism which go undetected or do not result in these outcomes are often not included in recidivism measures, leaving this research open to reporting biases. In addition, the type of offence can vary, and while the general category is often stated (e.g., sexual, violent, or general recidivism), the actual act and the severity of the offence is often missing. This means that reductions in severity or frequency of offending are unrecorded, we only have access to dichotomous information (i.e., recidivism or not). Finally, lower base rates and under-reporting observed for some types of offending (e.g., sexual abuse) make it difficult to link DRF changes with reductions in recidivism, particularly with shorter follow-up periods for detection. Despite these limitations, research linking DRF change with recidivism (as a proxy for changes in behaviour) is currently the closest available measure for the causal link between DRF, change, and offending.

2.5.1. Meta-analyses and reviews

One of the central works investigating the link between intra-individual change and recidivism is Serin et al.'s (2013) review. Serin et al. (2013) comment on the lack of high-quality research investigating the link between intra-individual change and recidivism. Despite this, support was identified for the link between changes in several major DRF and recidivism, and 27 per cent of possible effect sizes identified were significant. The authors interpret this

finding as support for the conclusion that there is a significant relationship between DRF score change and recidivism, and that this makes them both core risk factors and key targets of change (thereby supporting the need principle). However, while these findings are somewhat promising, the authors note that many of the putative DRF included within the studies did not demonstrate the expected relationships with change and recidivism. The authors conclude: "little is confidently known about which factors are the most productive targets for change, whether dynamic risk factors truly behave in a dynamic way to affect future recidivism, and if change predicts recidivism beyond knowledge of static risk" (p. 50). It is therefore clear that despite significant advancements being made in the empirical investigation of DRF, the evidence for the relationship between DRF change and recidivism is perhaps not as consistent as is commonly presumed. For this reason, the authors emphasise the need for significant advances in terms of how we conceptualise and measure these constructs.

More recently, van den Berg et al. (2018) conducted a meta-analysis exploring the predictive properties of dynamic risk assessment instruments frequently used with individuals who have sexually offended. The authors found that change scores were significantly related to recidivism only after controlling for static and initial dynamic risk scores, this relationship was small, and DRF score change only partially explained changes in rates of recidivism. Despite the small effect sizes, the authors suggest that the results provide support for Andrews and Bonta's (2010) risk and need principles.

2.5.2. Individual studies.

The findings of recent individual studies appear to be similarly mixed in their results. While several studies have found support for the link between DRF change and recidivism (Beggs & Grace, 2011; Cohen, Lowenkamp, & Van Benschoten, 2016; Coupland & Olver, 2018; de Vries Robbé, de Vogel, Douglas, Nijman, 2015; Olver, et al., 2018; Vose, Smith, & Cullen, 2013), many have failed to identify consistent support for the link between DRF change and recidivism, instead finding either mixed, weak, or limited support (Labrecque et al., 2014; Olver et al., 2014; Wakeling, Beech & Freemantle, 2013; Wooditch et al., 2014), support for only one factor (Kroner & Yessine, 2013), or no support at all (Howard & van Doorn, 2018; Mastromanno et al., 2018; Morgan, Kroner, Mills, Serna, & McDonald, 2013; Wilkinson, 2005). For example, Olver et al. (2014) found "very small to moderate" changes in DRF which were "weakly and inconsistently" related to recidivism (p. 544). When these authors controlled for baseline risk the link between treatment change and reduced recidivism often improved.

2.5.3. Summary: Change and recidivism

It is clear that there are a number of methodological issues affecting our ability to explore the relationship between DRF change and recidivism accurately (e.g., short follow-ups, measurement of recidivism, etc.). There appears to be evidence of a relationship between changes detected via repeated DRF measures and changes in rates of recidivism, yet this is not a strong or consistent relationship. Findings tend to be modest and researchers generally identify a number of methodological limitations. Overall it appears that the "evidence for a definite link between changes in dynamic risk factors and reduced recidivism remains limited" (Casey, 2016, p. 105).

2.6. The Role of Treatment in Dynamic Risk Factor Change

While change has been observed between pre- and post-treatment measures of DRF scores (and some with multiple time points), little analysis has been done to ensure that this is a direct result of treatment. For example, few studies investigate treatment integrity or dosage (e.g., Wooditch et al., 2014), which makes it difficult to identify what exactly may be changing during treatment. In addition, there appears to be limited understanding and theoretical attention to potential mechanisms of change (Grady et al., 2011; Kroner & Yessine, 2013; Serin et al., 2013).

My final consideration in this chapter is whether treatment facilitates observed changes in DRF and reductions in reoffending. This is an important question both theoretically in that it can help to determine the existence of causal risk factors, but also practically because it is the foundation of the need principle. If DRF change through treatment does not lead to a reduction in recidivism, then arguably we are squandering limited resources on a strategy that it not supporting the overall aims of the criminal justice system; namely reduction in harm to potential victims and, therefore, safer communities. Treatment outcome evaluations can help us to investigate this question, but only if they measure DRF change (i.e., proximal measures) rather than simply looking at whether an individual received treatment and if they reoffended. In addition, it is possible that DRF change may be linked with recidivism without treatment necessarily causing that change.

2.6.1. Meta-analyses and reviews.

Beggs (2010) completed a review investigating the link between within-treatment outcome and sexual recidivism and found that the evidence was fairly mixed. Until this time, little research had employed the use of proximal measures of treatment change, and so the amount of evidence available to draw from was limited. Three methods were surveyed in this

review: psychometric measures, risk tools, and clinical ratings. In the studies using psychometric measures, most effect sizes were non-significant, and some were significant in the wrong direction (i.e., positive changes were associated with higher rates of recidivism). Beggs (2010) concluded that "there is a lack of reliable and consistent findings linking within-treatment dynamic change (measured psychometrically) with decreases in recidivism, although some promising results have been reported" (p. 375). When looking at risk assessment tools measuring change, the results were also mixed, with the VRS: SO (sexual offending version) showing the most promise in terms of linking DRF change with recidivism. Reviewing systematic clinical ratings, Beggs (2010) found that some studies reported a significant relationship between clinician's ratings of positive change following treatment and reduced recidivism, while others did not.

Olver and Wong (2013) conducted a review of four treatment programmes for sexual offending which included looking at treatment change and recidivism. They concluded that there were mixed results concerning whether these programmes were able to reduce risk, but that "evidence of more compared to fewer treatment improvements were increasingly associated with more substantive reductions in recidivism outcomes" (Olver & Wong, 2013, p. 588). However, there was variation and the studies surveyed had their limitations; these authors suggest that "while offenders may make some important in-treatment psychological changes, these changes may or may not necessarily be risk related" (Olver & Wong, 2013, p. 588).

As mentioned in the previous section, Serin et al., (2013) conducted a review investigating the link between intra-individual DRF change and recidivism. When considering the role of treatment in the studies they reviewed, Serin et al., (2013) concluded:

> It is apparent within this review that therapeutic change does not consistently lead to reduced likelihood of future crime... Although it is especially difficult to defend programs when successfully adopting treatment skills does not translate into lower risk to reoffend, it is even difficult to defend successful programs when it is unclear which treatment elements are responsible for the change and which offenders have changed. (p. 50)

This conclusion supports the point made earlier, that we cannot be certain what is driving change and whether it is a result of treatment. It also suggests that we currently do not have enough evidence to support a causal relationship between DRF and offending.

Finally, a recent paper by Klepfisz et al. (2016) reviewed the evidence supporting the relationship between DRF for violence and recidivism. The authors found a lack of support for the notion that change in factors commonly considered DRF for violence are related to changes in recidivism – with only a few DRF displaying any relationship. Further, the authors emphasise that there is a lack of evidence that treatment itself reduces violence, and even less supporting the assumption that DRF change moderates this relationship (Klepfisz et al., 2016). This calls into question the legitimacy of targeting DRF in treatment, as is the recommendation of the need principle and thus considered best practice internationally.

2.6.2. Individual studies.

The above-mentioned inconsistencies are reflected in individual studies on treatment change and recidivism, where findings are mixed and there is little consensus. Some studies have found support for the notion that treatment is responsible for reductions in DRF scores, and that this change is linked to reduced recidivism (e.g., Beggs & Grace, 2011; Olver & Wong, 2011; Olver, Wong, Nicholaichuk, & Gordon, 2007). Others have found modest support for the role of treatment in DRF change and reductions in recidivism (Duwe & Rocque, 2016; Jung & Gulayets, 2011; Wooditch et al., 2014), and several others have demonstrated weak to no support (e.g., Bowen, Gilchrist, & Beech, 2008; Kroner & Yessine, 2013; Olver, et al., 2014; Wooditch, et al., 2014). More recently, researchers have demonstrated that treatment has been responsible for DRF score changes, but that these changes did not translate into changes in rates of recidivism (Barnett et al., 2013; Howard & van Doorn, 2018; Klepfisz, 2018; O'Brien & Daffern, 2017).

2.6.3. Summary: Treatment change and recidivism

As described above, there is inconclusive evidence that treatment is responsible for producing changes in DRF scores and whether this influences future behaviour. There have been several reasons put forward by researchers for why this might be the case. Firstly, it could be that treatment is not targeting the right variables in order to change DRF, or that DRF are not actually able to change in a way that reduces the likelihood that individuals will reoffend (Beggs, 2010). Alternatively, the measures of change utilised may not be valid (i.e., changes detected are not real); the fact that research using various methods has found different results supports this hypothesis (Beggs, 2010).

2.7. Conclusions: Empirical Evidence

This chapter has reviewed the recent empirical evidence concerning the relationship between DRF and offending. The range of DRF that are a part of routine correctional practice

are well established as strong empirical predictors of recidivism, and research continues to further understand these correlates. They are currently considered plausibly causal, and this assumption is reflected within the need principle of the RNR model, which states that these should be the targets of interventions aiming to reduce recidivism (Andrews & Bonta, 2010; Bonta & Andrews, 2017). However, numerous researchers have highlighted the importance of DRF meeting certain criteria if they are to be considered causal, for example changes in DRF should be linked with expected changes in criminal behaviour. Arguably DRF have satisfied some of these conditions, they have been shown to precede and predict reoffending (to varying degrees), but the extent to which *changes* in DRF are linked with changes in offending and the reason for any observed changes is at this stage uncertain. I have relied here on published studies which have measured DRF change over time, often in response to treatment, and then attempted to link these with reductions in recidivism. As discussed, there are numerous methodological challenges to the reliable and valid measurement of these constructs, and this is reflected in the variability in the quality of these studies and their findings. Indeed, if these were robust variables which demonstrate construct validity (i.e., measure what they claim to), we would expect to see strong evidence of a linear relationship with recidivism. This is not the case. While there are studies which have found that improvements in DRF (often through treatment) are statistically linked with reduced reoffending, there are others which have not, and further studies which may not have made it to publication due to null findings.

These discrepancies place the field in an unenviable position when it comes to attempting to change criminal behaviour. As pointed out by Nunes et al. (2019):

> It can be frustrating and demoralizing … to realize that the most common research designs used in studies on sexual offenders do not demonstrate, support, or suggest any clear conclusions about the causes of sexual offending or, consequently, about the factors that should be targeted in treatment. (p. 232)

It is apparent from my review of the literature that the assumption when researchers fail to find the expected evidence is often that this discrepancy is primarily an issue of measurement or operationalisation. Foundational issues such as construct validity are often overlooked, and findings are presented as a discrepancy from what should be expected (i.e., given other evidence). For instance, concluding that findings offer 'partial' support or are 'promising', despite the fact that the evidence found speaks more to the lack of association. In addition, if researchers find that static risk scores change they tend to attribute this to measurement error

or do not discuss this finding (e.g., Labrecque et al., 2014; Schlager & Pacheco, 2011), yet when DRF scores change, no such claims are made and it is presumed that they have changed as a result of treatment. This trend speaks to the strength of the assumption and investment in the notion that they are causally linked with offending, rather than being an intentional strategy to mislead readers. For example, the use of the word 'disappointing' when an expected result was not found (Kroner & Yessine, 2013). I am concerned that this tendency to look for confirmatory findings rather than acknowledging results which challenge the status of DRF may lead the field into a theoretical dead end (Ward & Beech, 2015). Perhaps most importantly, this will also have a detrimental impact upon our ability to design and implement interventions to reduce crime.

On the other hand, some empirical studies do note issues with their conceptualisation, as discussed above (e.g., Serin et al., 2013; van den Berg et al., 2018). Serin and Hanby (2016) comment that "the field has limited evidence of valid intra-individual change constructs and measures, necessitating the need for both theoretical advances and empirical research" (p. 1588). As this chapter has demonstrated, the evidence is limited and mixed, both in the quality of study design and methods, as well as their findings concerning DRF and offending. This suggests that as currently conceptualised DRF are not causally linked with offending, and furthermore, they could not possibly be. They are general, poorly defined constructs which contain a range of things (i.e., possible causes, contexts, states, behaviours), and these things overlap – DRF domains have no boundaries. If we take this stance, it is not surprising that these factors are not reliably linked with variation in offending behaviour.

There have been several recent suggestions for both empirical and theoretical research which may enhance our understanding of DRF, and thus help to guide interventions to ameliorate them. For example, van den Berg et al. (2018) suggest that (statistically) uncovering causal pathways between 'psychologically meaningful' risk factors and offending may provide more understanding of the constructs underlying risk factors. Another example is recent attempts to account for and address the complexity and heterogeneity of risk profiles (i.e., sets of problems) for individuals. For example, Taxman and Caudy (2015) used latent class analysis to uncover four risk profiles based on combinations of 'risks' and 'destabilisers'. While these suggestions and developments are admirable, they are unlikely to advance our understanding of the causes of (and therefore solutions to) crime, because of the conceptual problems with the DRF construct (discussed further in the following chapter). Sullivan (2019) suggests that in order to advance the study of DRF and crime researchers should engage in 'co-ordinated pluralism'. This involves carefully defining and operationalising crime related constructs and

facilitating interfield cooperation to uncover the mechanisms underpinning these constructs. Future research which aims to identify the relationships between these variables needs to first address the conceptual confusion and construct validity issues currently faced.

Finally, it has been suggested that "moving beyond a data-driven approach will require different ways to identify risk factors, and possibly in how we conceptualize them" (Cording et al., 2016, p. 90). This will involve advances in empirical methodology as well as *conceptual* and *theoretical* work – explored in the following two chapters. Indeed, "it is difficult to see why anyone would be interested in statistical associations or correlations if the findings were not in some way relevant to an understanding of causative mechanisms" (Rutter, 2007, p. 377). In my view, the empirically determined criteria discussed in this chapter are only one part of causality; if these criteria were met (which at this stage they are not) researchers could infer that there may be some common cause underpinning DRF and offending behaviour. However, researchers would still need to specify the mechanisms by which DRF and offending are causally related, and therefore how changes in these mechanisms may promote desistance. To take this further, I argue that advances in this research will involve *beginning* with conceptual tasks (i.e., determining valid targets of explanation), and then conducting empirical work to test the theorised relationships between mechanisms. In other words, before we can make advances in the measurement of DRF they must undergo a reconceptualisation; in their current form they are poor targets for measurement and explanation.

Chapter Three: Theory and Conceptualisation

The conceptualisation of dynamic risk factors (DRF) and the explanation of offending through the development of theory are fundamental concerns for forensic psychology. While statistical relationships are useful for informing prediction, they are not able to adequately guide our understanding of individual behaviour. It has been observed that "while a great deal is known about risk factors for offending, less is known about causes, or about causal pathways or mechanisms. Ideally, interventions should target causes of offending" (Farrington, Loeber, & Ttofi, 2012, p. 62). The simple reason for this is that "manipulating a predictor will not change the outcome, while manipulating a cause could" (Wikström & Treiber, 2017, p. 74). In practice, DRF are routinely treated as if they are causal (i.e., the need principle), reflecting a lack of attention to the difference between prediction and explanation. Appreciating the fact that correlation does not imply causation, it can be argued that the reliance on predictors in treatment and as the focus of theories is misguided. Perhaps a more fruitful way forward is to develop explanations of offending behaviour which can account for the frequent co-existence of DRF but do not rely on them as explanation in themselves. Before I discuss this possibility, it is necessary to understand the pervasiveness of the issue.

DRF are generally the foundation of explanations of offending; they were imported from the research domains of measurement and prediction into the generation and development of theory. This chapter will focus on existing theories of offending and outline several conceptual problems with DRF which make them poor candidates for causal explanation. I will begin by outlining the most influential theories of offending, all of which rely on DRF as central concepts. I will then put forward a number of conceptual criticisms of DRF and introduce several recent theoretical approaches which have aimed to overcome these limitations by shifting the focus of theories of offending. I will conclude this chapter with some comments about the potential role of DRF in explanations of offending, setting the scene for the following chapter, which takes a more in depth look at theory development.

3.1. Existing Theories of Offending

Theories of offending are most often based upon correlational analyses of the phenomena associated with offending (i.e., DRF); they are driven by 'ground level' data (Cording et al., 2016). While this makes sense as we want to be able to describe the functional relationships between variables, it is also problematic because correlation does not indicate causation in any straightforward sense. Theoretical and conceptual analysis is an important phase in the practice of scientific inquiry alongside the detection of phenomena and their

interrelationships (Haig, 2014; Ward & Beech, 2015). The focus of theories ranges from general explanations of criminal conduct (e.g., the PCC) to more specific behaviours. Interestingly, many of these theories are concerned with explaining sexual offending against children (some of these will be discussed in this chapter), but there are also specific theories developed for rape, violence, domestic violence, and so on.

There are typically three levels of theory identified in the field of forensic and correctional psychology, each dedicated to explaining different targets (Ward & Hudson, 1998a). The first level consists of multi-factorial theories that combine several factors or domains thought to influence offending behaviour. For example, Marshall and Barbaree's (1990) *integrated theory* and Ward and Siegert's (2002) *pathways model* are both intended to explain sexual offending, and encompass self-regulation, cognitive, affective, interpersonal, and sexual domains of risk. The second level of theory is concerned with explaining single factors and focuses in depth on specific categories or domains of risk, such as offence-related attitudes and cognitive distortions, or relationship difficulties. The third level is dedicated to descriptions of the offence and relapse process, such as Ward and Hudson's (1998b) *self-regulation model*, and Finkelhor's (1984) *pre-conditions* for sexual offending. More recent theories have aimed to integrate all three levels, or to provide causal explanations based on human functioning rather than focusing on correlates (discussed later in this chapter).

In addition to these theories which inform case formulation and treatment for (mainly sexual) offending, researchers have recognised the need for rehabilitation frameworks to guide practice more generally (i.e., to spell out how changes should be made to prevent crime). The most influential of these frameworks will be discussed in depth towards the end of this thesis when I discuss the implications of this research for practice. However, at this point it is worth noting that rehabilitation frameworks are underpinned by theories of offending. They contain over-arching values (e.g., harm reduction), principles and guidelines (e.g., the RNR principles), and assumptions about the causes of offending, which rely upon various theories (Ward, Melser & Yates, 2007). For example, the RNR model is grounded in a *general personality and cognitive social learning* (GPCSL) theory of criminal conduct.

3.1.1. A general personality and cognitive social learning theory.

Bonta and Andrews (2017) have developed a comprehensive model of crime based on their central eight risk factors, the GPCSL. This perspective brings together empirical research (i.e., meta-analyses and prediction) and theoretical work in order to inform practice within the criminal justice system. It is built upon criminological theories and, as the name suggests, views

social learning and reinforcement as particularly important in the initiation and maintenance of criminal conduct. The central eight risk factors are proposed to influence the cognitive decision to commit a crime within any given situation (Bonta & Andrews, 2017). For example, the presence of multiple risk factors (e.g., antisocial attitudes, offence-supportive associates, antisocial personality) increases the likelihood that an individual will commit an offence. The GPCSL acknowledges a range bio-psycho-social influences upon the development of criminogenic needs and criminal behaviour, for example family of origin, neighbourhood features, temperament, gender, age, ethnicity, etc.

While this model describes the functional relationships between DRF and criminal conduct, in its current form it is unable to explain the onset and/or reoccurrence of crime for three major reasons. First, because the DRF in the model are essentially *summaries* of putative causal factors, contextual features, and mental state variables, it is unclear exactly what structures and processes they are actually referring to. In other words, they are inherently vague and are not coherent theoretical constructs (Ward & Fortune, 2016a). Second, the GPCSL is a *functional* model that does not provide a description of the causal mechanisms constituting each DRF nor explain how they influence each other. Third, relatedly, there is no attempt in the model to describe how the different DRF interact to cause crime and its reoccurrence. In my view, the GPCSL is best construed as a descriptive, conceptual model that loosely links background factors and predictors (i.e., DRF) to crime. The PCC distinguishes between risk (correlates and predictive variables), need (dynamic and functional/causal variables), and strengths (risk reducing variables), but does not offer an adequate explanation for how these variables collectively determine offending behaviour. It lacks depth when it comes to specifying the mechanisms underpinning criminal behaviour (Fortune & Heffernan, 2019) as it is built upon its correlates rather than the causes of behaviour in general.

3.1.2. Theories of sexual offending.

In addition to this general theory of criminal conduct, various researchers have developed theories based upon specific DRF for sexual offending. I will now outline several of the most influential theories and point to some of the problems with these explanations of crime. It is worth noting here that these theories are all based primarily on empirically identified correlates of sexual recidivism and that these roughly converge in four (overlapping) domains of functioning: cognitive, interpersonal, self-regulation, and sexual.

Finkelhor (1984) proposed four *preconditions* which must be satisfied for a sexual offence to occur. The first is a motivator which predisposes an individual to engage in an

offence. This can be one of three risk factors: emotional congruence with children, sexual attraction to children, or blockage of appropriate means to achieve sexual needs (Finkelhor, 1984). The second precondition is overcoming internal inhibitions, for example by rationalising and justifying behaviour or using substances to temporarily suspend judgment. The third is overcoming external barriers such as gaining unsupervised access to a potential victim, and the fourth involves overcoming any resistance from the victim. This theory proposes that the first two preconditions are internal causes and the final two explain how a motivated individual engages with other people and their environment in order to facilitate an offence. While Finkelhor's (1984) preconditions integrate a number of observable features of sexual offences or individuals who have committed these (e.g., DRF such as cognitive distortions, self-regulation problems, emotional congruence, sexual deviance) and accounts for different motivations (e.g., sexual needs, intimacy), it lacks depth and clarity. For example, it is unclear what exactly is involved in emotional congruence, blockage, or rationalisation. In order to explain how an individual becomes motivated and then overcomes internal inhibitions theorists must refer to psychological processes and mechanisms.

Marshall and Barbaree's (1990) *integrated theory* incorporated both distal and proximal influences on the development of sexual offending. This theory sees childhood adversity causing the development of problems associated with sexual offending (i.e., DRF) through biological development and social learning. These problems then persist and result in the development of additional risk factors, for example problems with interpersonal functioning leading to social rejection and loneliness in adolescence. When in a context where there is opportunity (e.g.., unsupervised access to children) these vulnerabilities can lead to a sexual offence. The reinforcing nature of the offence and post-hoc rationalisations and justifications for behaviour maintain offending over time. While this theory represents substantial progress in the field, due to its broad nature it is overly vague and lacks depth. It can account for the development and maintenance of problematic sexual behaviour over time at a general level but cannot explain how these vulnerabilities may cause an offence at any time.

Hall and Hirschman's (1992) *quadripartite theory* combined cognitive distortions, affective dyscontrol, sexual arousal, and personality problems. Personality problems are seen as enduring vulnerabilities, while the other three factors depend on both internal and external states. The quadripartite theory proposes that one of these four factors functions as a 'primary motivator', while all four are likely implicated in an offence. The primary motivator is the strongest of the four risk factors and acts to push the individual over the threshold necessary to commit an offence. The reliance on one factor as driving the offending is somewhat limited,

but it does allow for the categorisation of individuals according to their primary motivator. For example, those who are driven by sexual motives versus those who offend to cope with emotional problems. This may be useful in terms of identifying general problem areas to target in intervention, however it lacks clarity in terms of the interaction between the four types of risk factor. In addition, the reliance on 'personality problems' as a catch all category for personal vulnerabilities is vague.

Ward and Siegert's (2002) *pathways model* built upon these existing theories, also suggesting four mechanisms; intimacy and social problems, distorted sexual scripts, cognitive distortions, and poor emotion-regulation. These authors also propose that a primary mechanism drives behaviour, but that all four domains are implicated in offending. For example, distorted (i.e., paedophilic) sexual scripts may lead to sexual offending when an individual is experiencing rejection from peers and is able to either justify or suspend judgment of the behaviour. This model represents progress as it acknowledges the heterogeneity of pathways to sexual offending and the interaction between DRF, but it is lacking in the sense that it still does not explain precisely *how* they interact to lead to an offence.

Beech and Ward's (2004) *etiological model of risk* further built upon the theories previously discussed and outlined the relationship between different types of risk factor. According to the etiological model, stable DRF are psychological dispositions which indicate a vulnerability to commit an offence and static factors, such as criminal history, are evidence of these traits manifesting in the past. Acute DRF arise when vulnerabilities are triggered (e.g., by peers, conflict, substance use) resulting in internal states (e.g., distress, intoxication) which lead to an offence. This distinction between traits and states is coherent with a propensities model of risk and the idea that psychologically meaningful risk factors or vulnerabilities may or may not lead to an offence. This model aimed to bridge the gap between empirical and theoretical research by attempting to explain how these predictors may over time be causally related to risk. However, like the previous theories it was built upon, it lacks depth and clarity. It offers a useful way to understand the overlap between different types of risk factor (i.e., intersectionality), but it does not possess adequate detail to explain how risk factors interact to cause crime.

The final theory I will mention here is Ward and Beech's (2006) *integrated theory of sexual offending* (ITSO). The ITSO widened the scope of etiological theories by incorporating biological and neurological influences on offending. It includes the four domains of risk discussed above and aims to increase understanding of their interaction through the concept of 'interlocking neuropsychological functions' (Ward & Beech, 2006). These interactions occur

between three systems; motivational and emotional, perception and memory, and action selection and control, and they cause the problems which lead to sexual offending (i.e., DRF). The ITSO is novel in its combination of developmental and evolutionary perspectives with proximal influences such as neurological mechanisms and contextual features. Unfortunately, due to its wide scope it is limited in its detail on any one aspect of the theory. In addition, it overlooks the first-person experiential level of understanding, meaning that it does not adequately account for human agency and self-determination. This is a shortcoming which more recent theories have aimed to address (see below).

The theories discussed here are centred upon combinations of empirically supported DRF for sexual offending. For example, they combine problems such as sexual deviance, cognitive distortions, interpersonal difficulties, and poor self-regulation. They aim to explain how these problems can lead to sexual offending, but they tend to overlook critical sources of motivation and agency, and overall lack depth in being able to explain why persons have the problems exhibited and how exactly they lead to offending for some individuals but not others. I believe that due to their focus on DRF and specific outcomes (i.e., illegal sexual behaviour) they have inadvertently ignored important determinants of behaviour. They are incomplete explanations because they focus on negative outcomes and correlates but do not include normative or adaptive aspects of human functioning, those psychological mechanisms which influence human behaviour more generally. For example, motivation, values, attributions, expectations, etc. I will now outline numerous conceptual problems with the DRF construct which make them unsuitable targets for the explanation of crime, before discussing more recent theoretical developments which aim to remedy these issues.

3.2. Conceptual Problems

There are a number of conceptual problems with DRF which undermine their putative causal status. The first problem is that DRF *lack coherence*; they are *composite constructs* which contain several different types of variable (Ward & Fortune, 2016a). While they likely contain causal strands, in their standard form they are more like general categories that also incorporate contextual (e.g., gang membership), behavioural (e.g., watching child pornography), and psychological state aspects (e.g., feeling lonely). Secondly, due to the co-existence of these composite categories and their multiple potential causal strands, DRF *lack specificity*. That is, they are unable to identify which potential cause is relevant for explaining certain phenomena. Third, DRF *lack precision* and suffer from the *grain problem*, which means that there is little agreement concerning which level of abstraction is the appropriate one to

interpret them at. They are often formulated at various levels: as general or umbrella categories (e.g., antisocial personality) or more fine-grained categories composed of specific features (e.g., impulsivity, hostility). Fourth, DRF *lack factualness* because they are not 'scientific kinds' (Ward, 2016). They are (at least in part) normative constructs, which only exist due to their co-occurrence with behaviours and outcomes which society has deemed harmful and/or unlawful.

In order to make these conceptual problems more concrete, I will apply them to the example of 'lack of emotionally intimate relationships with adults' or 'intimacy deficits'. This DRF category has strong empirical support for its association with sexual recidivism and relies upon contextual evidence such as having no stable partner relationship currently or in the past, or having intimate relationships characterised by conflict and infidelity (Mann et al., 2010). It encompasses those who desire and those who avoid intimacy, and it has been acknowledged that "these varying facets of dysfunctional intimacy may have different underlying pathologies and so may lead to different treatment targets" (Mann et al., 2010, p. 201). It is further suggested that intimacy deficits could be an indicator of other DRF, such as sexual deviance, attachment problems, and poor emotion-management (Mann et al., 2010). Thus, it is unclear what exactly the term refers to and how it might increase risk at the individual level (i.e., it *lacks specificity*). In addition, it is a *normative* category in the sense that intimacy levels vary across the population and are only deemed to be at a dysfunctional level when they co-occur with more serious norm violations such as sexual offending.

In order to illustrate the remaining two problems, *lack of coherence* and *lack of precision* or *grain problem*, I will explore the composite nature of the intimacy deficits category. The cluster of factors which hang together across explanatory levels and grains of analysis (i.e., from general categories to specific processes) to cause or constitute this DRF include (but are not limited to):

- *Cultural/contextual level*: norms specifying the types of relationships that are appropriate and what these should involve; gender norms; ideal sources and amount of intimacy (e.g., what are intimacy *deficits*?); laws (e.g., age of consent); and social opportunities for connection.
- *Interpersonal/social level*: interpersonal skills (e.g., communication); social learning; support/advice, social roles; social isolation; expectations and responsibilities.

- *Phenomenological/psychological level*: emotional connection/congruence; empathy; sexual preferences; beliefs about relationships (i.e., self and others); perspective-taking; emotion-management; and attachment style.

- *Neuropsychological level*: brain regions and neurotransmitters, such as oxytocin, vasopressin, pre-frontal cortex, and hormones that underpin psychological problems and experiences indicative of intimacy problems.

- *Biological level*: biological sex; sexual arousal; physical health; and physical attributes (i.e., size, attractiveness).

Each level of description relies upon various sources of evidence and varies in its level of abstraction. The umbrella category 'intimacy deficits' encompasses more specific features at lower levels, such as emotional (e.g., congruence with children) and cognitive (e.g., beliefs about children and sex) processes. No level on its own can provide an adequate explanation of intimacy deficits, as unique properties emerge across levels. These examples are by no means exhaustive, but hopefully illustrate the range of influences evident in just one DRF category, and also the significance of overlap between different categories (e.g., intimacy, cognition, emotion, and sexual). In addition, the inclusion of different types of constructs within a single DRF category is incoherent and creates confusion. This example provides support for the assertion that DRF categories are of little use for the purposes of explanation. They are acceptable predictors, but if they are to explain (re)offending or be imported into the treatment domain (via case formulation) they need to be reworked. While it is likely that the DRF domains identified contain causal elements, due to their vague definitions it is impossible to tell what potential cause is being referred to. For example, it is unclear whether social isolation (one aspect of the intimacy deficits category) is caused by over-sensitivity to interpersonal threat, shyness, a lack of social skills, environmental conditions, or any combination of these and other interpersonal influences.

3.2.1. Conceptual problems with protective factors.

Fortune and Ward (2017) identified four similar problems with PF: definitional ambiguity, explanatory confusion, their dual status, and practitioner uncertainty. *Definitional ambiguity* results from inconsistencies in the way PF are conceptualised across studies and the initial importation of the concept of 'protection' from the child maltreatment literature. This means that the object of protection has shifted from the child experiencing maltreatment to potential victims, and thus protection runs the risk of being equated with risk reduction (Fortune

& Ward, 2017). *Explanatory confusion* results from a lack of theoretical attention to PF and the way they may function to reduce risk – this relates to their composite nature, vagueness, and the lack of explanatory depth provided by PF categories. The *duality* of these concepts refers to the problem that PF and DRF are largely variations in the same phenomena. For example, what might be considered risky for one context or individual may be a strength for another (i.e., social skills can be used for prosocial or antisocial means). This conceptual relationship is unclear, and thus broad category labels cannot offer causal explanations for differences in offending behaviour. The final issue is *practitioner uncertainty*, or confusion about the status of PF in practice as correlates, causes, or social facilitators of desistance (Fortune & Ward, 2017). Practitioners (and indeed researchers) lack a sophisticated understanding of how PF relate to DRF conceptually and within individuals. These mirror the conceptual problems with DRF; that they are normative rather than scientific or natural kinds, they contain a range of attributes, behaviours, and circumstances, and there is a lack of understanding about how these concepts may be theoretically related.

While researchers such as Polaschek (2016, 2017), Thornton (2016), and de Vries Robbé et al. (2015) have explored the relationships between DRF and PF, this work is arguably weakened by its incorrect presumption that these constructs are theoretically coherent. If you accept that they are best characterised as general labels for clusters of predictive items, then strictly speaking they cannot interact with each other or exert any causal influence on the world. The same argument holds for discussions about whether they are unipolar, opposite ends on a continuum and so on; they are statistical tools and as such have no causal implications for the way psychological and social processes determine crime and its related problems. In essence, my argument is that DRF and PF are constituted by heterogeneous conceptual elements (e.g., possible causes, contextual factors, and mental state attributes) and lack causal specificity (i.e., it is not clear which out of a range of possible causes a particular DRF or PF represents). While heterogeneity can be considered an asset in a predictive construct, it is a major weakness in explanatory ones, and rapidly leads to problems with conceptual coherence and lack of clarity about what psychological and social processes are being referred to. Ward and Beech (2015) take up this point and argue:

> What follows from the argument that DRF are composite constructs rather than coherent theoretical concepts developed within the context of an explanation of offending? The first implication is that they do not pick out psychological processes at all but rather are labels or summaries of multiple factors... In a

strict sense, they do not really refer to anything real; they do not exist. Second, the various types of dynamic risk factors (e.g. antisocial personality, relationship problems, intimacy deficits, self-management problems, emotional dysregulation, and distorted attitudes…) are markers for psychological and social causal processes, aetiological factors and symptoms or clinical phenomena...On their own they only function as predictors of reoffending; but as such they arguably track causal processes and to that extent provide a valuable function. (p. 7)

To summarise, DRF and PF are broad, composite categories that lack specificity, contain multiple (sometimes conflicting) concepts, and are overly general (Fortune & Ward, 2017). Each construct is comprised of a range of internal and external characteristics, including underlying propensities and observable manifestations of these within various contexts. The second fundamental issue is that the features and processes that DRF and PF refer to are defined by their statistical relationship to a range of (positive or negative) outcomes (i.e., reoffending or desistance) within large groups. This means that not only are they normative in the sense that they are defined in terms of their relationship to harmful and illegal behaviours (i.e., they are not scientific, natural, or 'real' categories), but also that they point to trends across populations rather than causes of individual behaviour. In this sense they are predictive rather than explanatory tools. They do not refer to genuine causes of crime and thus cannot point to theoretically or personally meaningful targets for correctional treatment. I have also touched upon the dual nature of DRF and PF as contextually bound; they are risky or protective based on their influence upon individuals within certain contexts rather than being fixed causes (Fortune & Ward, 2017). These observations cast significant doubt upon the idea that DRF and PF can interact in order to influence individual behaviour and undermine our assumptions about the nature of risk. This further suggests that we are mistaken in our near exclusive focus on correlates of criminal behaviour, and our tendency to categorise these as DRF or PF. In my view, it is more likely that these categories represent variation or impairments in functional domains that support or enable goal-directed behaviour in general, for this reason I now turn to human agency.

3.3. Recent Developments: Human Agency

Recent theoretical developments have aimed to address the conceptual issues with DRF by developing theories based on agency and goal-directed functioning (e.g., Heffernan &

Ward, 2015, 2017; Serin, Chadwick, & Lloyd, 2016; Thornton, 2016). From this perspective behaviour is motivated by particular needs or desires, and then a set of psychological mechanisms or capacities (e.g., attention, memory, control, schemas, etc.) allows individuals to engage with their environment in goal-directed action. When applying this perspective to criminal behaviour, DRF are aspects of persons and their environments which *indicate* a higher probability that these goal-directed actions will involve crime. For example, obstacles which get in the way of goal-directed behaviour (i.e., problems with self-regulation, control, problem-solving, substance use, environmental constraints) or aspects of a person's behaviours and environments which indicate a pattern of criminal engagement (i.e., associates, history, attitudes toward crime). Thus, DRF are symptom-like features which can be measured to estimate risk but cannot themselves be causal (although they may reflect the manifestation of underlying causal mechanisms).

3.3.1. Agency model of risk.

The *agency model of risk* (AMR; Heffernan & Ward, 2015) is a general model of goal-directed behaviour which was developed in response to the above problems with DRF. It is a dynamic interactional model which depicts the relationship between an agent and their context and proposes that both play a part in facilitating behaviour. The agent is comprised of three 'levels' which give rise to the development of goals and action sequences. These levels are biological systems, social roles, and personal identity. Depending on contextual triggers and internal motivators (i.e., unmet needs) one or more of these levels may influence the development of a goal. For example, a teacher (social role) that views himself as vulnerable and misunderstood (identity) and is experiencing rejection (social context) and loneliness (i.e., interpersonal needs unmet) may form the goal of sexual intimacy in general, or perhaps because attempts with women have previously failed the goal is specifically sexual intimacy with a student. Due to his unsupervised access to children (physical and social context) and his understanding that sexual contact with minors is illegal (normative context), he selects a child and identifies strategies to gain their trust. He then implements these plans and monitors their success, altering his strategies and goals where necessary in response to his context. The AMR is a useful preliminary model because it represents a shift in focus to the capacities and processes underpinning agency, but it lacks depth in that it does not unpack the role of cognition and emotion across action sequences.

3.3.2. Thornton's theories of dynamic risk and protective factors.

Thornton (2016) recently developed a theory of DRF based on the *good lives model* (GLM; Ward & Maruna, 2007), *schema modes* (Beck, 1996), and the *theory of reasoned action* (TRA; Fishbein & Ajzen, 2010). He argues that the dominant view of DRF as offence supportive propensities neglects human agency and so fails to focus on the causes of criminal behaviour. The GLM is utilised as a framework to understand sources of motivation (i.e., fundamental human interests); essentially answering the question: what is it that people seek? The second element of this theory involves goal-directed decision-making; how do people guide their behaviour to meet these interests? The TRA offers an account of reasoning based on outcome expectancies, social pressure, and self-efficacy. Essentially the attractiveness of behaviour, how others are likely to judge it, and how successful one expects to be in enacting a behaviour. Schemas are involved as an intermediate zone between goals and reasoned action, containing representations of environmental features such as people, places, and situations (e.g., scripts and strategies). These representations allow individuals to weigh up the potential success and attractiveness of various types of action, and they can be adaptive or maladaptive in nature.

In applying this same theory to desistance processes and items from the SAPROF (de Vogel et al., 2009), Thornton et al. (2017) developed a theory of PF. At an abstract level, Thornton et al. (2017) propose that PF can be understood within a broader context of DRF; the two categories are viewed as "families of related concepts" (p. 30). It makes sense that Thornton's (2016) theory of DRF applies also to PF, and that goal-directed processes that lead an individual towards prosocial behaviour rather than offending are linked with PF. At an intermediate level, Thornton et al. (2017) propose that many PF may be conceptualised as arenas within which certain protective processes occur (e.g., employment, relationships). Finally, they state that PF will take different forms and operate in different arenas depending on various characteristics of persons and their specific concerns. Four PF categories are proposed in the theory: *dynamic internal*, *social*, *professionally provided*, and *openness* to *professionally provided* PF. The examples used (i.e., SAPROF items; de Vogel et al., 2009) are a mixture of internal abilities and characteristics, and the arenas within which they operate. For example, dynamic internal factors (e.g., empathy, coping skills, and self-control) may facilitate engagement with social contexts (e.g., work, leisure, relationships) and openness to professionally provided factors such as supervision and treatment. These examples mirror internal and external risk domains, and in this sense can point to surface-level aspects of internal and external desistance processes (instead of offending), but do not offer a more in-

depth theoretical account of how PF reduce risk. Despite this limitation, the idea that DRF and PF are category labels for two groups of inversely related characteristics and the arenas that they manifest within is a simple but attractive one, as there has long been confusion regarding the theoretical relationship between risk and protection.

3.3.3. Durrant's theories of dynamic risk and protective factors.

Durrant (2017) presents an evolutionary perspective on DRF and PF, based upon individual differences and environmental conditions. His framework contains two models outlining how and why developmental PF reduce the likelihood of future criminal behaviour (Durrant, 2017). The first is based on the evolutionary theory of *life history strategies* (Belsky, Schlomer, & Ellis, 2012) which vary in their pace depending on individuals and environmental features. For example, a fast life history strategy typically develops when an individual is exposed to environments which are dangerous and unpredictable, as this is adaptive in terms of reproduction and survival. In contrast, a slow life history strategy develops when an individual exists within safer or more predictable environments, as there is less need to protect the self and others from threat; the individual engages in long-term thinking, as there is no reason to speed up the reproduction process. From an evolutionary perspective, the point of an adaptation is to increase an individual's chances of surviving in order to pass on copies of his or her genes to future generations. If organisms evaluate their life situation as risky then it makes sense from a strictly biological viewpoint to engage in reproduction at an earlier age before they suffer injury or even death. A faster strategy is associated with greater risk taking, violence, and reckless or impulsive behaviour, while a slower strategy is associated with a more balanced lifestyle and well thought out plans. It should come as no surprise that Durrant (2017) links slower life history strategies with developmental PF, and faster life history strategies with DRF. This makes sense from an evolutionary perspective, and suggests that perhaps offending, DRF, and related behaviours can be considered adaptive in some contexts; that is to say that they can be useful in self-preservation, reproduction, and survival – at least in the short term.

Durrant's (2017) second model is based on the evolutionary concept of *plasticity*, or differences in an individual's susceptibility to environmental conditions. Unpacking the concept of plasticity, Durrant (2017) states that individuals are able to:

> …reliably adjust their behaviour in response to enduring features of the environment. However, there is also a widespread recognition among evolutionary biologists that there are also significant individual differences in

plasticity that reflect the outcome of selective processes. In other words, individuals vary in the extent that their behaviour is plastic in response to features of the developing environment. (p. 8)

He proposes that whether a developmental feature is a DRF or PF (i.e., functions to increase or decrease risk) depends upon how easily the individual is influenced by their environment. In this sense, PF are positive environmental features *and* their influence on the individual. For example, a stable home environment may be protective, but individuals will differ in their sensitivity to this developmental PF (or opposite DRF). It has been suggested that those who are more easily influenced by contexts may be more sensitive to both positive (PF) and negative (DRF) conditions, and vice versa. However, in some cases individuals may be more or less sensitive to PF than DRF. Thus, plasticity may offer some explanation as to why some individuals with risk factors do not offend, and also why certain people may be less responsive to positive external influences (e.g., social control, professional support).

Durrant (2017) concludes that "before we can make theoretical progress in understanding antisocial behaviour … and the factors that predict those individual differences (risk and protective), we need to have a general understanding of – for want of a better term – our 'human nature'" (p. 6), and that "in order to understand antisocial behaviour we need to locate such acts within the broader context of normative human development" (p. 6). These suggestions are in line with my aim to provide a theory of human functioning that can explain antisocial behaviour, generate and guide future theoretical work, and ultimately inform more effective forensic practice.

3.3.4. Agency filter model

Serin et al.'s (2016) *agency filter model* is another recent attempt at providing an explanation of the relationship between DRF and PF and offending. It was developed in response to the confusion surrounding what PF are and how they might ameliorate risk. This model proposes that DRF and PF exert their influence upon behaviour via a filter containing "agency, self-identification, hope, optimism, self-efficacy, attributions, social supports" (Serin et al., 2016, p. 164). This filter represents the agent's interpretation of events and situations which could be risky or protective; essentially how they choose to respond to (mainly contextual) PF and DRF. For example, whether attaining employment is meaningful enough to direct a person away from offending. The idea of an individualised filter can begin to explain why the same factors may influence individuals differently, and how other 'non-criminogenic'

factors such as identity and self-efficacy may influence risk of recidivism. It brings agency and self-determination into the equation but arguably lacks depth in terms of explaining how these factors might interact to cause behaviour. In addition, it seems that DRF and PF in this model mainly consist of external events and situations or internal states which the agent responds to. It pays less attention to an individual's capacities and does not account for the role of dysfunctional psychological mechanisms (e.g., impaired self-control) in some instances of offending.

3.4. Conclusions: Theories of Offending

The theoretical progress made in the field of risk prediction and management is substantial. Empirical researchers and theorists have gone from listing the correlates of reoffending to developing a psychology of criminal conduct and theories describing one or more DRF in detail. This work has informed correctional and forensic assessment, management, and treatment in a way that is consistent with evidence rather than being based upon subjective judgments. Within forensic and correctional practice DRF are best viewed as predictive devices that are generally effective in assigning probabilities to individuals based on group membership (e.g., risk bands such as low, medium, or high), but on their own are unable to tell us what might influence individual risk. I argue that the relationship between predictors and causes (i.e., treatment targets) is unclear, and that the role of DRF in treatment depends on their ability to explain why individuals offend and what might facilitate desistance.

In my view DRF are best viewed as boundary riders; they tell us there are a number of possible problems operating within a particular domain such as crime supportive beliefs, contextual features, mental states, social situations, and psychological dispositions. But they do not tell us what the specific problems are. For that we need to enlist explanatory theories and move beyond 'one size fits all' approaches to assessment and treatment. The following chapter will further discuss the development of explanatory theories and the potential role of DRF within explanations of offending. Chapter four acts as a bridge between this evaluation of DRF (and PF) and my suggestions for moving the field forwards towards more comprehensive and useful explanations of behaviour.

Chapter Four: Causality, Values, and Theory Development

As explored in the previous two chapters, there are empirical and conceptual requirements to consider when inferring a causal relationship between variables such as DRF and outcomes such as recidivism. I will begin this chapter by exploring a set of causal criteria put forward by Bradford-Hill (1965) and, drawing from the conclusions of the last two chapters, summarise the ability of DRF to satisfy these requirements. Then, in line with the normative nature of crime and DRF discussed earlier, I will explore the role of *values* in forensic psychology and scientific inquiry (Heffernan & Ward, 2019), before discussing the generation, development, and evaluation of theory. I will then outline some recent suggestions for theory development which are utilised throughout the remainder of this thesis. This includes careful consideration of the targets of explanation moving forward, attention to mechanisms underpinning phenomena, multi-disciplinary research, and use of integrative pluralism in linking theories. Towards the end of this chapter I will suggest a shift in focus from the correlates of offending (i.e., DRF) to the types of goal-directed practices which they manifest within and the capacities underpinning human agency (Heffernan & Ward, 2019).

4.1. Causal Criteria

Before I address the important topic of theory development, it is worthwhile to briefly consider the extent to which DRF in their current form are useful targets for the explanation of crime. Their central role within explanations depends upon the extent to which they can be considered causally related to offending. In order to explore this, I will use a set of causal criteria from the epidemiological literature. Bradford-Hill (1965) put forward nine criteria which can assist researchers in determining causal relationships, which are:

1) Strong *statistical association* with a specific outcome
2) *Consistency* (i.e.., across places, circumstances, time, and observers)
3) *Specificity* (i.e., to particular groups, body systems and sites, and diseases)
4) *Temporality* (i.e., a putative cause precedes an outcome)
5) *Biological gradient* (i.e., a decrease in effect with a decrease in cause)
6) *Plausibility* (i.e., is the cause reasonable within the context of current knowledge? Ideally an etiological/causal mechanism should be identified)
7) *Coherence* (i.e., does it cohere with knowledge about the domain? Established facts act as epistemic constraints on causal inference)

8) *Experimental* manipulation (i.e., evidence from well-designed studies supports a cause and effect relationship)

9) *Analogy* (i.e., the existence of other similar causal relationships)

The extent to which DRF meet these criteria has been somewhat addressed in the previous two chapters. To summarise, the ability of DRF to predict recidivism at a level better than chance suggests that there is a statistical *association* between these constructs and recidivism. This is likely to be stronger for some DRF than others. The *consistency* across contexts and raters has also been demonstrated to some degree, however, their cross-cultural utility has been seriously questioned (Shepherd & Lewis-Fernandez, 2016), and inter-rater reliability varies by factor, for example, "items requiring greater subjective judgment, such as marital/family factors, use of leisure time, and peer relationships, have significantly lower reported rates of reliability" (Baird, 2009, p. 7). Thus, criteria one and two have been partially met, but I would argue not adequately.

The *specificity* of DRF to both individuals who have committed offences and offending contexts is also unsupported. For example, relationship issues, alcohol abuse, and poor self-regulation or impulsivity are experienced by individuals in the general (i.e., largely non-offending) population, indeed even specific factors like paedophilia may be observed in individuals who have not engaged in sexually abusive behaviour. However, offence-related attitudes and emotions and use of illegal substances may be more specific to offending. This is partially due to their normative nature and the fact that they are defined in relation to the criterion (i.e., they are offence-supportive or offences themselves). In light of these normative issues, I argue that criteria three has not been met.

In regard to criteria four, longitudinal studies have demonstrated that DRF *temporally* precede recidivism. However, the extent to which they were present prior to the initial offence has not been established for all DRF. Therefore, perhaps some DRF function as maintaining factors of criminal lifestyles rather than initial causes of offending. For example, an individual may develop errors in thinking or gravitate towards antisocial peers following incarceration, but perhaps they were not present during the initial offence. Arguably, if the outcome is recidivism rather than the initial onset of offending, then DRF meet criteria four. The *biological gradient* (i.e., increases and decreases in recidivism) and *experimental manipulation* (i.e., through treatment) of DRF were explored in chapter two, and the relationship was not found to be linear; therefore, criteria five and eight are not adequately supported.

In terms of the *plausibility* of DRF as causes of offending, the description of functional relationships between DRF and offending (i.e., the PCC) seems to give a superficial account of their relationship. However, as discussed in the previous chapter, it falls short of specifying etiological or causal mechanisms. In addition, some DRF are easier to link with crime than others, for instance antisocial attitudes and personality are more plausible causes of crime than unemployment and relationship issues (i.e., the latter may be symptomatic of these problems, or be more normative and related to behaviours other than crime). Therefore, criteria six has not been fully met, DRF require further specification in order to be considered plausible causes. For example, in order to make relationship issues or interpersonal problems a plausible cause of violent behaviour we would need to identify what the problem is (e.g., communication, expectations, jealousy or fear of abandonment, etc.) and how it is linked with violence (i.e., the function of violence such as control or emotion expression, whether this process was planned or impulsive). This level of specificity requires theories which refer to psychological mechanisms and processes.

Further, as discussed above, DRF are themselves incoherent constructs, which undermines their ability to be considered *coherent* with existing knowledge about offending. However, they may be considered coherent with the broader knowledge about the domain (i.e., the PCC and other theories of crime), the reason for this being that they are often the basis of these theories. Therefore, criterion seven, coherence, may be partially met, but this may be a side effect of the reliance on these correlates as explanations for crime. Finally, the criterion of *analogy* (criterion nine) or the existence of similar causal relationships has not been met. For instance, there is a lack of mechanistic or causal explanations linking psychological phenomena and behaviours in related disciplines such as psychopathology (i.e., the debate surrounding psychological symptoms and disorders; discussed further below).

In conclusion, while DRF may temporally precede and be statistically associated with recidivism (to a degree), they fail to demonstrate the expected relationships and fall short of being plausible causes and coherent explanations for crime. One of the reasons for this is their value-laden nature; definitions of DRF and crime rely on social norms and expectations. For this reason, I will now turn my attention to the role of values in scientific inquiry. Part of the process of theory generation (and development) is the determination of suitable targets for these explanations, and as discussed in the previous chapter, the current targets (DRF) are value-laden concepts. I suggest that in order to progress our understanding of the causes of offending we need to shift our focus away from DRF and towards behaviour more generally, only then can we understand how DRF and offending are related.

4.2. Values in Forensic Psychology

According to Sadler (2005), two notable features about values are: 1) they are action guiding in the sense that they provide *reasons* for action and can be translated into specific goals and plans, and 2) norms reflecting values are used to *evaluate* actions, persons, and outcomes as worthwhile or unworthy. The particular nature of the *worthiness* depends on the type of value in question, and the specific context and set of practices concerned (Tappolet & Rossi, 2016). Thus, values are prescriptive in that they communicate to individuals that they 'ought' to evaluate certain things highly or 'ought not' to do so. Statements containing values can be more or less widely endorsed but are not normally viewed as true or false. However, they may be considered objective in the sense that actions based in values can be the result of an impartial (relatively unbiased) inquiry process agreed to by independent decision makers (Douglas, 2009). For example, well-designed research and interventions that save lives or reduce harm (while not causing undue damage) are considered valuable. Johnson (2014) highlights the interactional nature of values, arguing that "some state of affair is valuable for or to some organism, animal, or person. …. *nothing is valuable in itself* …. but only in relation to how it serves a living purposive organism or group of organisms" (p. 49; italics in the original). The field of forensic psychology exists because it is widely considered valuable to understand the causes of offending and thereby reduce the social and financial costs associated with crime (i.e., via prevention and/or intervention).

In the early 1970s, the field of correctional rehabilitation encountered a crisis of confidence, due in part to Martinson's (1974) essay suggesting that perhaps "nothing works" to reduce recidivism. This work prompted Canadian psychologists James Bonta, Donald Andrews, and colleagues to conduct meta-analyses of treatment outcome studies to empirically determine "What Works" in reducing reoffending. Scientifically oriented rehabilitation frameworks such as the RNR model (Andrews & Bonta, 2010; Bonta & Andrews, 2017) have guided the development and evaluation of numerous correctional programs for a multitude of offence types including sexual and general violence. The PCC and RNR model were born out of a need for (and the value of) evidence, hence their 'rational empirical' orientation and prioritisation of prediction and data over explanation and theory development. There are reasonable grounds for accepting that programmes constructed in line with the RNR principles, including a focus on altering criminogenic needs (DRF), are currently the most likely to be successful in reducing recidivism rates (Andrews & Bonta, 2010). Thus, adherence to the relevant correlates of offending seems to be the best way to guarantee successful outcomes. From this perspective, failure to follow the scientific evidence when designing policies and

intervention programmes is unethical and irrational; scientific inquiry is the best way to discover the causes of offending and this is essentially an empirical process. The policy of grounding criminal justice practice in scientifically warranted evidence is no longer seriously contested by forensic or correctional practitioners (Bonta & Andrews, 2017; Gannon & Ward, 2014; Taxman, 2017). However, as outlined in previous chapters, the focus on correlates and empirical criteria for establishing causality is challenged by numerous researchers.

While it is appreciated that there is a normative aspect to forensic and correctional research and practice, this is often viewed as somewhat external to the activity of science. It does not and should not directly influence the day-to-day operation of scientific inquiry; that is, values are not internal to science in this domain. In a nutshell, the mantra is: follow the evidence and keep values out of the picture. They are subjective, ideological, and are likely to result in derailment of good research and ultimately what Andrews and Bonta (2010) have termed "knowledge destruction". While this is a simplification of the viewpoints of empirically oriented researchers in the criminal justice domain, it is close enough. There is an emphasis on detecting factors related to offending and identifying the functional relationships between these factors and subsequent outcomes. The science underpinning forensic and correctional practice is hardnosed, factually based, and value free in its central activities. In fact, in the subject index of the PCC (Andrews & Bonta, 2010), there are no entries at all under the heading of "value" and only one under that of "norm", and this refers to the definition of crime rather than norms as the values central to forensic research and practice. Although there is some discussion about the values of diversity, autonomy, and collaboration on pages five to eight (Andrews & Bonta, 2010; Bonta & Andrews, 2017), this is relatively peripheral to the elaboration of the RNR model. Values appear to be regarded as an add-on rather than as fundamental to every aspect of inquiry and correctional practice.

In my view, the relationship between scientific practice and values is more complex and far reaching than is typically depicted. It is generally acknowledged that the criminal justice arena is essentially a normative one characterised by a number of foundational values such as punishment, deterrence, law, responsibility, guilt, remorse, accountability, harm, redemption, and so on. However, less discussion has focused on the ways that values influence forensic and correctional research and practice. For example, values are directly and indirectly related to the practice of science and are integral to its structure and procedures. Values determine what is researched (i.e., social and ethical values and concerns; e.g., sexual abuse, violence), and how studies are designed and their results interpreted. For example, epistemic or cognitive values determine what is considered good research (e.g., objectivity, validity, reliability, bias, control).

In addition, every treatment programme and intervention model presupposes a specific conception of what constitutes a meaningful and worthwhile life, and relatedly, what is considered a successful outcome (Day & Casey, 2009; Day & Ward, 2010). The targets of intervention, typically DRF, PF, mental disorders, and psychological problems, refer to both factual conditions and valued/disvalued states of affairs; normative elements are built into their meaning. Finally, human beings are motivated by a number of natural and socially acquired needs and interests, all of which contain normative components by virtue of indicating possible future harms or benefits.

Failure to appreciate the pervasiveness of values in the generation of knowledge does not mean that they are not influencing research, but simply that it is unacknowledged. A danger is that certain types of knowledge or evidence are privileged over others or that problematic assumptions remain unquestioned due to theoretical and ideological allegiances which distort the detection and explanation of phenomena.

4.3. Theory Generation, Development, and Evaluation

According to Haig (2014) science typically proceeds as follows: constrained by a developing problem comprising a set of empirical, conceptual, and methodological considerations, certain data are brought to the researcher's attention and are ordered via the detection of one or more phenomena. For example, in the forensic field researchers gather data including behavioural observations (i.e., offending history), self-report (i.e., descriptions of and reasons for offending), and information from risk assessment or psychometric measures. This information is typically used to infer the existence of various phenomena which have been deemed relevant for predicting offending, for example, attitudes, impulsivity, interpersonal problems, deviance, etc. Once detected, these phenomena are explained via the generation of theories, for example, abductively inferring the existence of a causal mechanism underpinning 'impulsive' violence. Here, *abductive inference* involves reasoning from a presumed effect (i.e., the phenomenon) to its explanation in terms of an underlying causal mechanism (i.e., the theory). In the forensic field this means that DRF (as the most robust phenomena inferred from correlational data) are the focus of explanations of offending. Because there are patterns of behaviour which *often* co-exist (e.g., substance abuse and recidivism, sexual deviance and sexual abuse) explanations often rely upon these regularities as functional explanations and fail to consider that these co-existing phenomena are underpinned or caused by various mechanisms. For example, the PCC proposes that criminogenic needs increase the likelihood

of recidivism, without specifying the mechanisms through which they influence the decision to offend (Andrews & Bonta, 2010).

Theory development is an ongoing evaluative process which brings researchers closer to the truth, and therefore provides the scientific community (and other related fields) with useful explanations on which they can base predictions and interventions. From an initial judgment of the plausibility of an explanatory theory, ideally attempts are made to elaborate on the nature of underlying mechanisms, frequently by way of constructing plausible models (i.e., adding explanatory depth to the theory). When the theory is well developed, it is evaluated against a number of criteria including its empirical adequacy and criteria principally to do with the explanatory worth of the theory. For example, the theories outlined in the previous chapter were focused primarily on DRF, and so they suffered from the same conceptual problems as these constructs (e.g., incoherence, the grain problem). Thus, *epistemic or cognitive values* have very important roles in theory development (Douglas, 2014; Haig, 2014) as well as in empirical studies. Researchers make decisions about which theories to pursue and typically will appeal to values such as fertility, falsifiability, predictive precision, explanatory depth and scope, and external consistency (Ward, Polaschek & Beech, 2006); these are the features which indicate a theory's potential, and in a sense they represent promissory notes. Minimally, all theories should be logically consistent and be able to account for the available empirical evidence, otherwise they are not scientific theories at all (see Douglas, 2014).

Once existing theories have been evaluated and their limitations identified, researchers can strive to develop better explanations – for example using techniques such as *integrative pluralism* or *analogical reasoning* to address the relevant issues and capitalise on the strengths of existing theories. These and other suggestions for developing theory will be discussed later in this chapter. At this point it is worth mentioning the importance of methodological frameworks in guiding ongoing theory development, particularly in the realm of forensic psychology where the central constructs suffer from the conceptual problems discussed above. I suggest that the field of forensic and correctional psychology has encountered a 'dead end' when it comes to theory development (Ward, 2019; Ward & Beech, 2016). This situation can only be remedied by developing new methods of theory development and expanding the focus from DRF and offending.

4.3.1. Methodological frameworks in research.

A general methodological framework offers a way of unifying the diverse range of cognitive tasks involved in reaching descriptive and explanatory goals. The value of adopting

a theory of scientific method is that it unifies the diverse tasks that constitute research within an overarching general framework. These tasks include formulating an initial question, designing a study, choosing methods for collecting data, analysing data, detecting explanatory targets (i.e., phenomena), inferring etiological and compositional factors (i.e., the structures and processes constituting mechanisms that underlie phenomena), and so on. Without a methodological framework it is easy to get lost in the research process and run the risk of squandering limited cognitive, social, and financial resources. Indeed, Thomas and Sharp (2019) partially attribute the lack of mechanistic explanations in psychopathology to "the absence of a metatheoretical framework that defines what mechanisms are and how they may be pursued" (p. 3).

Consistent with this view, it makes sense to construct general methods of inquiry to guide researchers in their attempts to identify and explain crime and its related phenomena. If we accept that DRF track causal processes in some way, then coming up with a general method to help isolate and model potential causes at multiple levels of analysis will be invaluable. This kind of methodological framework will be nested within the type of general scientific method discussed above; it should provide guidance in picking out promising causal factors from the research and clinical literature, and in constructing rich descriptive models of their constituents and subsequent causal impact. The *risk causality method* (RCM) outlined in chapter six of this thesis is an example of such a methodological framework which has built upon a number of suggestions from recent literature.

4.4. Suggestions for Theory Development

At this stage it is useful to consider five suggestions for theory development put forward by Ward (2014). Two of these suggestions follow the arguments made thus far: 1) that researchers need to prioritise uncovering the causal mechanisms underpinning phenomena of interest, and 2) that individuals' values, beliefs, and experiences be considered worthy of analysis. The additional three suggestions will be discussed further in this section: 3) that the targets of inquiry should be clearly identified from the outset, for example through the construction of an *exemplar* or typical instance of the phenomena of concern, 4) that research should *build upon existing theoretical work* across disciplines rather than reinventing the wheel, and 5) that researchers utilise *integrative pluralism* to bring together a range of explanations spanning various levels of abstraction and analysis.

4.4.1. Targets of explanation and classification.

An important decision in any comprehensive research project involving the formulation of theories concerns the nature of the explanatory targets; what phenomena have been detected and how do we explain them? In the criminal justice context, it makes sense that the focus of explanation should be the behaviour that we wish to change, although it is acknowledged that 'crime' is a broad and value-laden category containing various behaviours and outcomes. Since there are a set of correlates that are commonly relied upon as indicators that predict future behaviour, any comprehensive explanation of crime or desistance should also be able to explain the frequent but varied co-occurrence of DRF and PF. There are numerous examples of these behaviours and characteristics contained within etiological theories, lists of correlates such as the central eight (Andrews & Bonta, 2010), assessment tools, treatment manuals, and literature from related areas such as criminology and developmental psychology. The reason DRF and PF are included as explanatory targets alongside offending behaviour is that some of them (e.g., attitudes, propensities, abilities) are currently the closest thing we have to 'psychologically meaningful' features of offending (Mann et al., 2010). However, I emphasise the importance of explaining offending behaviour over its correlates, as DRF and PF do not always accompany offending or desistance in typical ways, and indeed many of these characteristics may be present in the general population – they are human attributes independent of crime.

In more developed sciences the classification of objects like chemical elements or animals is based on the idea of 'natural kinds' (Magnus, 2012). According to Magnus (2012) a natural kind is a category within a domain of inquiry that facilitates "inductive and explanatory success" (p. 48). Inductive success means that once scientists have placed an entity within a category, they are able to infer the existence of further properties. For example, once you know that Poppy is a rabbit, you can infer that she is warm blooded, has fur, is likely to breed at a high rate, will dig burrows given the opportunity, has a certain life span, likes particular foods, and so on. The category of 'rabbit' and its associated causal properties also figures in explanations of Poppy's behaviour; she is innately wired to dig burrows in order to create a safe place to have offspring, thereby increasing their and her own survival chances. Natural kinds map onto causal properties and features of objects in the world and, by virtue of this relationship, can be appealed to in explanatory theories. Objects that share clusters of properties are grouped together and ordered within a scientific classification system.

Unfortunately, this logic of classification breaks down in the criminal justice arena and therefore, the categories used do not accurately map onto real objects or processes in the world. For example, if you know that Tom is a 'sexual offender', you cannot reliably infer anything

other than that he is a person who has engaged in an illegal behaviour of a sexual nature which has resulted in a conviction. Even though it may be more likely that he has deviant sexual interests (compared to the general population), this is not necessarily the case – there is no DRF with a strong enough relationship with recidivism that its presence could be inferred simply through an individual's classification by offence type. Offence categories such as sexual or violent refer (loosely) to types of behaviour, not coexisting clusters of properties (i.e., their proposed causes and associated 'symptoms'). Legal categories are closely aligned with social and ethical values, and they group together specific types of normative violations based on the harm they cause. Because offence categories do not pick out natural co-occurring psychological and behavioural properties, but rather reflect social and ethical normative decisions, they are not valid targets of explanation. Classification of individuals via the presence of DRF suffers from the same problems due to their normative nature. They are hybrid or composite categories typically composed of clusters of varying problems that do not 'hang together' in any meaningful sense.

Therefore, explanations aimed at different offence types and DRF lack specificity and arguably miss the mark. DRF are illusive, moving targets because they do not exist as kinds outside of normative conventions. The lack of validity of offence categories probably explains why there is considerable heterogeneity within each, and why criminal versatility and comorbidity of other problems (e.g., addiction, psychopathology, etc.) are prevalent. The reason is not necessarily because people who commit crimes are generally antisocial, but rather because the category of 'offender' is so loose that it fails to ground useful inductive inferences and explanations. Therefore, one challenge facing researchers in the criminal justice area is that the classification systems they are asked to work with are ones constructed from normative concerns. They do not map onto natural clusters of properties found in people who commit crime, and so research based on existing categories may have little to say about the psychological and social causes of certain types of norm violations.

Ward and Carter (2019) recently suggested that explanations would be better directed towards the motivational or functional systems which underpin certain types of practices (e.g., emotion management, mate seeking), rather than the outcomes or behaviours themselves (e.g., violence, sexual abuse). This is because the same action may have a range of reasons or causes, and different behaviours may be caused by the same underlying mechanisms. For example, problems with social interactions may be partially caused by an overly active threat detection mechanism, but may manifest within social isolation, social phobia, violence, sexual offending, substance abuse, etc. The way these problems manifest will depend on the functioning of other

related mechanisms, as well as environmental opportunities and constraints. One way to overcome this issue is to shift the focus of explanation to behaviours more generally, and then to apply these explanations to various criminal behaviours and their co-existing problems (i.e., explaining why features such as intimacy problems and sexual deviance *often* accompany sexual offending). In other words, while in the criminal justice domain we are concerned with preventing particular types of norm violation (i.e., offence types and associated DRF), it is important that explanations draw upon the mechanisms underpinning human functioning in general. These explanations are more likely to tap into natural kinds or real features of persons and behaviour (as opposed to 'offenders' and offences).

A potential way to do this, while respecting the importance of social and ethical values in identifying research problems, is to formulate the focus of inquiry in the following way. First, acknowledge the value of explaining why individuals intentionally inflict serious harm on others, such as sexual or physical violence. Second, create *exemplars* of offence types based on descriptions of the actions, and social and psychological characteristics of the typical individuals committing the offence (see Ward & Beech, 2015). Talking about mental disorders, Murphy (2016) states:

> An exemplar is a representation of the typical course and symptoms of a mental illness, whereas a model is a representation of those symptoms, that course, and the causal determinants of both of them. A model is an exemplar together with an explanation. (p. 206)

For example, in the case of sexual violence, one exemplar might describe an individual with longstanding sexual preferences for children, who has engaged in sexual interactions with children over a long period of time, emotionally identifies with children, has social and intimacy deficits, and believes that children are competent sexual agents (Mann et al., 2010). If the goal is prevention of sexual abuse, there should be multiple exemplars created, each designed to capture a prototype of someone who commits sexual offences (i.e., one exemplar cannot account for every person who commits a sexual offence). For example, capturing varying offence pathways, problem clusters, symptoms, and motivations. These should better capture the complexity of individual presentations, while also reflecting common processes.

Table 1

Domains of Risk and Protection and their Associated Practices[2]

Domain	Risk Factors	Protective	Practices, e.g.,	Capacities, e.g.,	Values, e.g.,
Sexual	Preoccupation Deviant arousal/sexual interest	Moderate intensity sexual drive Sexual preference consenting adults Attitudes support respectful age-appropriate sexual relationships	Sexual acts, emotional intimacy and communication, seeking sex, frequency/number of partners	Sexual drive, arousal, attraction (preference), acceptance of sexual identity, sexual scripts/schema	Pleasure, reproduction, safety, connection, mastery, inner-peace, fidelity
Interpersonal – separated into intimacy and peers or friendship	Emotional congruence with children Lack of emotional intimacy adults Lack of concern for others Negative social influences	Preference for intimacy adults Capacity for lasting emotional intimacy with adults Secure attachment Care and concern for others Social network External control	Partner choice, communication, emotional connection, establishing and maintaining bonds Peer choice, interests, influence	Interpersonal skills, preferences for intimacy, capacity for sexual and emotional connection Interpersonal skills	Pleasure, connection, relatedness, fairness, honesty and fidelity, harmony, support, mastery Support, friendship
Self-regulation	Lifestyle impulsivity General self-regulation problems Poor cognitive problem-solving	Goal-directed living Self-control Accept authority, rules Effective problem-solving skills In-tact cognitive functioning Intelligence	Seeking employment, managing finances, leisure activities, self-care, compliance with rules/law, problem solving, planning and goal setting	Motivation, self-control, skills relevant to context (e.g., attitudes, conflict resolution)	Autonomy/agency, mastery/success, creativity, safety/stability, contribution
Emotion management	Dysfunctional coping Grievance, hostility	Functional coping	Identification, tolerance, and communication of emotions, coping	Coping strategies, communication, emotion recognition and interpretation, control	Inner peace, comfort, pleasure, health
Attitudes	Offence-supportive attitudes Machiavellianism Hostility toward women Resistance to rules/supervision	Attitudes support respectful age appropriate sexual relationships Recognise others' rights Adaptive schema Trustful/forgiving Positive attitudes toward women Motivated and optimistic attitude toward desistance	Representing reality, causal reasoning, interpretation, attribution, explaining and justifying action	Memory, causal reasoning theory of mind, flexibility, interpreting input, accuracy (i.e., based on evidence)	Knowledge accuracy/utility, predictability, creativity, mastery/success

Table 1 contains the range of DRF and PF for sexual offending in the first three columns

on the left, exemplars should contain various combinations of these phenomena. It is

---

[2] *Note.* Reprinted from "Dynamic Risk Factors, Protective Factors, and Value-Laden Practices," by R. Heffernan and T. Ward, 2019, *Psychiatry, Psychology & Law, 26*, p. 316. Copyright 2018 by The Australian and New Zealand Association of Psychiatry, Psychology and Law. Reprinted with permission.

acknowledged that there are other ways to categorise and label these, but due to significant overlap and heterogeneity within categories, I believe this to be the most useful way to link DRF and PF with behaviour. Each domain contains a number of specific factors which are grouped together under an umbrella term. These domains are consistent with assessment tools and literature concerning sexual offending, but the specific wording and examples are mainly borrowed from recent work concerning the DRF with the most empirical support, and those considered most likely to be causes of offending (Mann et al., 2010). The set of PF are based on de Vries Robbé et al.'s (2015) list of plausible PF for sexual offending. Due to the lack of empirical research on PF, their list consists of the healthy poles of psychologically meaningful DRF (Mann et al., 2010), factors related to desistance, and PF assessment tool items (i.e., statistically linked with reductions in risk). They present their list of PF within eight domains, which are able to be loosely linked with the Mann et al. (2010) DRF, and which I have collapsed into the five risk domains in Table 1.

While DRF and PF suffer from numerous conceptual problems, they are included in Table 1 due to their widespread use and the fact that they are empirically linked with sexual offending. It may be useful from this point on to consider DRF and PF as broad categories that contain a mixture of descriptions of markers (i.e., predict more or less likelihood of reoffending), and rudimentary explanations that appeal to personal characteristics and contexts that could *plausibly* cause some of the variation observed in offending trajectories. In other words, DRF and PF contain underlying capacities (i.e., putative causes), their manifestation (i.e., practices), and contexts which are *associated with* increases or decreases in rates of reoffending. I suggest that the different types of goal-directed behaviours or practices (i.e., motivational systems directed towards normative outcomes) and the values and capacities underpinning them (the three columns on the right hand side of Table 1) are more useful targets to guide the initial construction of explanations. Sexual offending (and associated DRF/PF) is of interest to researchers due to the harm it causes, but the goal-directed behaviours themselves and the bio-psycho-social processes underpinning them are the central targets of explanation. These ideas will be explored in the following section of this chapter, where I will explore the ability of a 'practices' perspective to address the normative and value-laden nature of offending, DRF, and PF. My argument is that before we can reason about the possible functions of DRF and PF, we need to add depth to our understanding and explanations of these concepts by appealing to human capacities and psychological mechanisms. In addition, we need to understand these within the context of the values, norms, and practices that they refer to.

4.4.2. Developing mechanistic explanations.

The development of mechanistic explanations is increasingly being viewed as an important task for theory development; Ward and Fortune (2016a) note that:

It is not enough to simply state that there is a significant relationship between A and B. To be confident of a causal relationship and to be able to control and explain the outcomes, it is important to know how this occurs. Mechanistic explanations serve an important function by making the interactions between causal factors and phenomena much clearer. (p. 80)

Mechanisms are defined as structures which perform certain functions due to their parts and processes, they are dynamic and integrated (Bechtel, 2008). They are comprised of entities which interact to produce phenomena of interest, such as perception, memory, and other cognitive processes involved in action selection and control. These processes underpin goal-directed behaviours and so explain why we might see problems within certain types of practices, for example poor coping strategies or impulsivity.

As outlined above, the first important task of any theoretical endeavour is to clearly describe the targets to be explained. Hopefully it is clear by this point that DRF are unable to function as causes of offending. Instead they are better thought of as phenomena (based on patterns of data) which contribute to an exemplar – they are part of what needs to be explained rather than being causes which can explain offending. In line with the arguments above, the key targets here are the *causal mechanisms generating serious norm violations*. In other words, the focus is shifted from offending behaviour to the causal mechanisms responsible for norm violations such as problems with intimacy and sexual practices, and more generally self-regulation practices (e.g., emotion-management, self-control). Developing causal explanations for robust phenomena (i.e., patterns in the data such as the frequent coexistence of DRF indicators and recidivism) requires deeper analysis of the mechanisms which underpin certain mental states and behaviours. Once this is achieved, researchers can use exemplars to test the model or demonstrate how it can account for patterns of phenomena which regularly occur.

4.4.3. Utilising relevant theories: A multi-disciplinary approach.

Individuals who commit crimes exist and act within social and physical environments in order to attain valued outcomes. Therefore, it makes sense to utilise existing theory based upon our evolving understanding of persons and the ways in which they navigate their

environments. As above, the categories of 'criminal behaviour' or more specifically 'sexual offending' contain a wide range of actions and outcomes, enacted in various contexts and for diverse reasons. Likewise, the category 'offender' is not a natural kind and therefore does not generate useful inferences about the causes of behaviour. A view of those who offend as primarily persons requires the openness of researchers to look to fields outside of forensic psychology and the "What Works" literature in order to inform their understanding of both the causes of behaviour and mechanisms of change. It also requires the ongoing co-ordination of researchers across fields in order to form coherent and comprehensive explanations based upon comparable results (Sullivan, 2019). Sullivan (2019) observes that "investigators working within the same and across different fields of criminology have different ways of identifying, conceptualizing and classifying criminal behaviours and their causes as well as different methods and evidential standards for investigating them" and that these "serve as barriers to effective interdisciplinary communication that is necessary to facilitate causal discovery" (p. 562). This partially explains mixed findings within the empirical literature, and why existing theories have failed to progress our understanding of the causes of crime. Sullivan (2019) suggests a way forward through an interdisciplinary approach, where researchers across fields collaborate and are influenced by research from other disciplines.

Numerous related disciplines offer theories of human behaviour and psychological functioning based upon robust empirical findings. Evolutionary psychology is an important place to start, as a view of humans as evolved beings who are sensitive to environmental conditions has the potential to partially explain individual differences between those embedded in crime supportive lifestyles and those with a more prosocial orientation. In addition, the fields of criminology and forensic psychology will clearly have much to offer any explanation of crime, but only when viewed within a wider understanding of human nature (Durrant & Ward, 2015). In other words, what we know about persons who commit crimes is useful, but it is important not to overlook the fact that they share many characteristics with individuals who do not offend. Indeed, in some cases they may have more in common with other groups of non-offending individuals (e.g., those suffering from addiction or psychological disorders) than they do with others who have committed offences. Co-ordinated efforts by researchers skilled in different research methods can advance our understanding of the relationship between predictive constructs and recidivism, and help us to understand how they relate to other fields of psychology (Sullivan, 2019). However, this requires that researchers first come to a consensus regarding the definitions and operationalisation of DRF in order to be clear about

findings across studies (and disciplines) and the limitations of any inferences which can be made (Sullivan, 2019).

Putting the issues with DRF conceptualisation aside for now, I propose that any framework that explains criminal behaviour should also be able – in principle – to explain *all other forms* of goal-directed behaviour. In other words, it should be a map of all determinants of any behaviour, including motivation, decision-making and associated cognitive processes, emotion, and biological and environmental influences – in this context applied to our explanatory targets (i.e., various combinations of the contents of Table 1). Paralleling recent approaches to explaining psychopathology, norm violations such as crime are best understood as primarily arising from dysfunctional psychological systems *or* normal functioning systems operating within maladaptive environments (e.g., loss of behavioural control versus intact control within contexts that reward antisocial choices). This requires a multi-level, multi-factorial, and multi-disciplinary theory. This does not mean that one theory must explain everything, indeed it is unlikely that a theory could be entirely comprehensive without losing explanatory depth or becoming overly complicated. It requires integrative pluralism, or the linking of theories in friendly coalition.

4.4.4. Integrative pluralism and levels of explanation.

Integrative pluralism (Mitchell, 2003) is the methodological process of linking *local theories* across multiple levels of explanation in order to form comprehensive theoretical structures that can account for a larger range of phenomena. Local theories might focus solely on one psychological mechanism or level of explanation, and thus offer a partial explanation for a phenomenon which can be enriched by being linked with other theories. For example, the way that we explain our own behaviour (a subjective first-person perspective) differs from the way that psychological (or biological or sociological) theory would explain our decision-making processes, interpersonal functioning, positive and negative emotional states and traits, and so on. Each perspective offers a unique explanatory style which cannot be reduced or eliminated (i.e., important information is lost when social interactions are reduced to neurological processes). For the simple reason that a first-person perspective is the subjective point from which we engage in goal-directed behaviour, I propose that this is an important and irreducible level of explanation (Baker, 2015; Ward, 2014, 2016). However, which level is prioritised in any given explanation will depend on the task at hand. For example, in therapy the subjective first-person level of explanation might be most useful, however, when considering the use of medication, neurological and biological levels are more relevant. In

addition, social and contextual explanations may assist decision-makers in managing individuals within the community and in environmental strategies to prevent crime.

Keeping in mind the dual nature of DRF and PF, researchers should be meticulous in making a distinction between the search for factors that *cause* particular outcomes and the *normative* judgments that these outcomes are harmful or beneficial, and in what respects. This points to the importance of explicitly thinking about different levels of explanation, those concerning psychological (and biological/neurological) mechanisms versus norms and practices. One way to do this is to adopt a 'levels of analysis' framework such as the *research domain criteria* (RDoC) project from the field of psychopathology classification (Cuthbert & Kozak, 2013; Lilienfeld, 2014; Ward & Fortune, 2016b) to guide research into the causal mechanisms generating behaviour. Looking to analogue fields such as psychopathology, which is currently facing significant challenges similar to those experienced in forensic psychology (e.g., searching for causal mechanisms, comorbidity, transdiagnostic criteria, etc.), can offer potential avenues for future research.

Adapting the RDoC framework for the forensic and correctional areas, the emphasis would initially be on core psychological domains of functioning such as working memory, self-regulation processes, or negative valance systems (e.g., threat detection) and investigating how these may cause crime related problems in certain contexts (Ward & Fortune, 2016b). Therefore, the initial locus of explanation would be sub-personal mechanisms underpinning individuals' actions, such as the processes and components comprising working memory or emotion regulation. For example, the discovery that depletion of serotonin levels can result in lowered mood, which in turn might lead to the increased salience of norms associated with sex and intimacy seeking. When investigating these domains of functioning researchers can speak about dysfunction of mechanisms (i.e., instances where they do not function as they should, for example attentional biases, chronic low mood, etc.). These are likely to influence behaviour and, in some cases, may be relevant to offending.

The social, ethical, and prudential norms (i.e., personal goals) that direct the practices constituting DRF and PF in everyday life can then be studied separately, in light of the causal information yielded by empirical research into causal mechanisms. Once research attention shifts from sub-personal mechanisms to particular actions constituting DRF such as emotional dysregulation, researchers must consider goal-directed practices, which are constituted by norms and their associated actions. For example, an explanation of child sexual abuse that invokes desire for (and value of) sexual intimacy, rather than lowered serotonin levels. Norm-based explanations are typically formulated in terms of reasons, and reasons track values and

beliefs. These will be explored in the following section of this chapter, where I suggest shifting the focus towards goal-directed practices.

In line with this shift, in the following chapter of this thesis I will provide a *provisional* model within which a set of local theories (i.e., concerned with explaining a specific set of phenomena at a particular level of analysis – genetic, molecular, neural networks, phenomenological, behavioural, etc.) may be linked to account for crime and its related problems. Later I will point to areas for further development but not provide these local theories myself; this is a substantial task beyond the scope of this thesis. Each *capacity* contained within the provisional framework comprises one area of human functioning, and thus relates more or less to certain normative practices and domains of risk (e.g., relationship issues, poor self-regulation, and anti-social attitudes; see Table 1). However, because it is difficult to separate these internal capacities from each other and from environmental conditions, there will be significant overlap and the eventual linking of local theories is crucial. Each local theory should focus on one or more of the *subcomponents* and *processes* within each capacity, or across capacities. These local theories will include multiple levels of explanation, for example, behavioural, phenomenological (first-person experiential), psychological, neurological, biological, social, and contextual. These local theories provide compositional explanations of mechanisms and as such will increase the explanatory depth of the general model of human functioning (agency) presented in the following chapter.

4.5. Shifting the Focus: Goal-directed Practices

In line with the above suggestions, I now propose a shift in focus from risky characteristics, behaviours, and contexts (i.e., DRF), to the kinds of *practices* (i.e., goal-directed actions) these descriptions refer to. Embedded within these practices are values (i.e., priorities, motivators, norms), and underpinning them are human capacities. Identification of these capacities and relevant contexts (i.e., norms, opportunities) can inform rehabilitation which strengthens or alters them to support healthier and less harmful functioning. I will now explore the extent to which DRF and PF are value-laden and discuss what this means for theory development. I begin by outlining the influence of values and norms upon practices and subsequent judgments about the degree to which they manifest in offending behaviour. I will then use case vignettes (i.e., exemplars) to explore two DRF and PF domains (interpersonal and self-regulation) with respect to relevant normative practices and the values which influence them. This will demonstrate how a shift in focus from DRF to goal-directed practices can account for the observed problems (i.e., DRF) alongside more normative or adaptive cognitive processes, offering a more comprehensive account of the determinants of behaviour. I will also

make some preliminary suggestions regarding the sorts of capacities involved in these practices (i.e., those which facilitate healthy functioning).

4.5.1. Values and human agency.

From a naturalistic viewpoint, normativity (i.e., the existence of values) is pervasive in the natural world as well as in human culture (Thompson, 2007). In fact, social and cultural values are thought to have their origins in the basic biological needs and physical conditions that enhance individuals' chances of surviving and possibly flourishing (Johnson, 2014). Values refer to those aspects of the world that confer benefits to organisms and alert them to the presence of threats or possibility of harm. The influence of values is evident in the norms that govern the functioning of different action sequences in animals such as predator behaviour, workings of biological systems, the application of human moral systems, and primate social relationships. In speaking of the ubiquity and essential roles of norms in biological and social systems, Christensen (2012) views "normativity as inherent in the organization or form of living systems, specifically in the form that generates their unity and hence explains their existence" (p. 104).

Normative principles are natural in the sense that they specify the functional parameters of biological systems and social practices and have their origin in the development of agency in organisms of all types. Norms are reflected in goals, and the strategies selected to further these goals are evaluated against these norms, typically in a fluid, dynamic, and immediate manner, in 'real time.' In complex animals such as human beings, the capacity to flexibly adjust goal-directed strategies and plans in response to changing environmental contingencies is in part due to cognitive capacity and the availability of social and cultural resources (Sterelny, 2012). As Damasio (2010) comments:

> I see value as indelibly tied to need, and need as tied to life. The valuations we establish in everyday social and cultural activities have a direct or indirect connection to homeostasis. That connection explains why human brain circuitry has been so extravagantly dedicated to the prediction of gains and losses, not to mention the promotion of gains and the fear of losses. It explains, in other words, the human obsession with assignation of value.…Value relates directly or indirectly to survival. In the case of humans in particular, value also relates to the quality of that survival in the form of well-being. (pp. 47-48)

The link between values and action regulation is an important one for forensic and correctional researchers and practitioners. It underlines the need to understand offending behaviour within a relational model, taking into account the goals and interests of agents alongside the degree to which the social and physical environment actively supports norm violations. It also reminds us that goals and values only gain motivational power against an affective backdrop, they are conduits within which emotion and their associated appraisals influence the environment.

4.5.2. Goal-directed practices.

In the context of everyday activities, values are embedded within *practices*. Practices are the application of practical knowledge within goal-directed action sequences, governed by "a structured body of norms" (Wallace, 2009, p. 11). Norms are evaluative in nature and spell out whether or not an activity is done properly; if it meets the socially accepted relevant standards. For example, the practices constituting 'intimacy' refer to *objects* of intimacy or attraction (i.e., types of people), the *contexts* they exist within (e.g., time, place, social group), and the *behaviours* enacted (e.g., communication, sex). The normative commitments inherent within intimacy practices spell out what is widely considered desirable, acceptable, and healthy behaviour in intimate relationships. Because of their focus on *successful* action, practices have both *causal* and *normative* dimensions. Practices are goal-directed cognitive and behavioural activities that are intended to address specific tasks such as problem solving, planning, explaining and justifying action, establishing and maintaining relationships, regulating emotions, engaging in sexual activity, and so on.

Practices are underpinned by causal mechanisms and the capacities they constitute. For example, a set of causal mechanisms might create difficulties in one's capacity to infer mental states in another person, resulting in harmful intimacy-related practices (e.g., sexual offending) in certain contexts (e.g., when experiencing loneliness and/or intoxication). If a crime is committed, then the hypothesised causal condition(s) or practice(s) involved are given the status of DRF. For example, the mechanisms thought to underpin intimacy problems, or the practice of substance abuse are only DRF if retrospectively observed to precede a criminal offence (i.e., an illegal practice). This means that DRF can come in and out of existence contingent on changes in the law or in ethical standards. Thus, what was once thought to be a DRF may cease to be one (e.g., changeable factors predicting homosexuality), and what was once considered to be benign or even beneficial might now be viewed as harmful (e.g., sex with early adolescents 300 years ago was socially acceptable whereas now it is a crime). This fact underscores their partly normative status, and thus their reliance on values.

I would like to stress here that the argument that DRF, PF, and their manifestation (i.e., illegal practices) are value-laden does not mean that scientific inquiry plays no role in their meaning (or in the referents of these concepts). They are hybrid constructs that contain *both* factual components (i.e., observations) and normative ones (i.e., value-based judgments). The purpose of examining the degree to which they are value-laden is to point out that normative discussion is an indispensable part of correctional research and practice. Figure 1 depicts a range of influences upon professional judgments concerning risk and highlights the crucial role of values and their associated normative commitments both in shaping and evaluating practices. A danger for forensic practitioners lies in focusing risk detection and subsequent treatment upon enduring personal capacities (often in combination with risky contexts) *and* adopting the default assumption that this is a value free process, guided solely by facts and their evidence. In reality, values (i.e., personal and professional sets of normative commitments) not only shape professional judgments, but they also provide the context within which offences occur. We are unable to remove values from the process of assessment and treatment, because values of various types (e.g., ethical, social, empirical, prudential) are present within the practices that we are trying to change (e.g., reoffending, desistance), as well as within the practices and objectives of the criminal justice system as a whole.
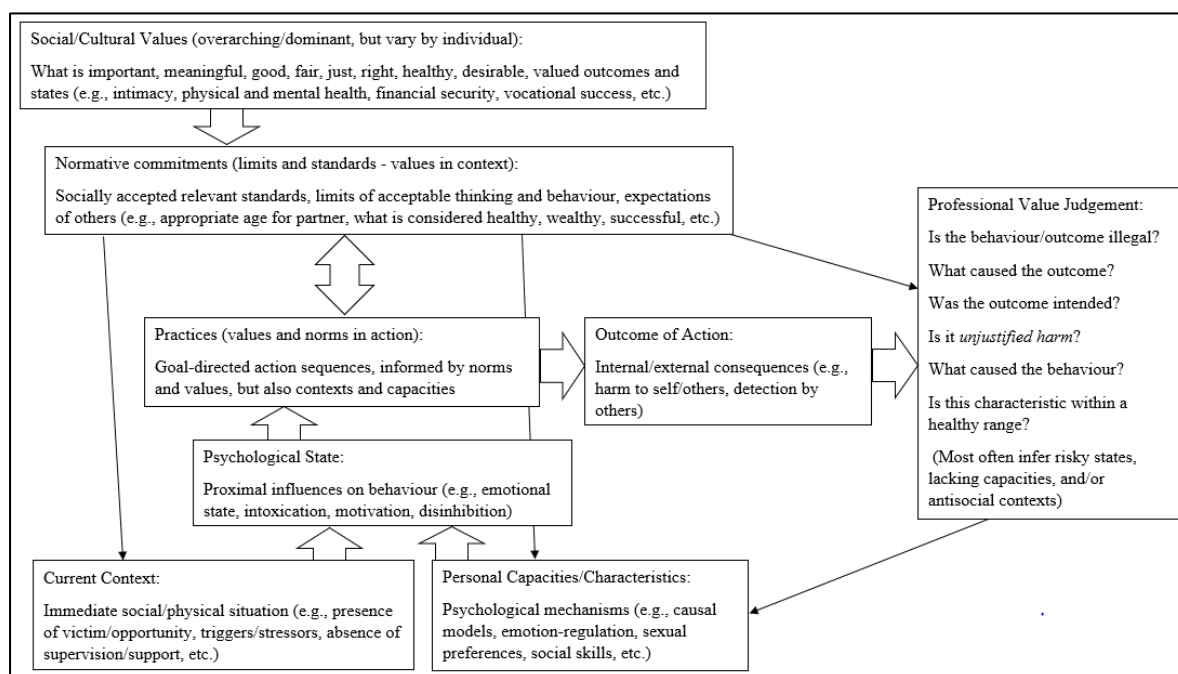

Figure 1. The role of values and norms in professional judgments of illegal practices. [3]

When conducting an assessment for the purpose of sentence planning, informing treatment, or managing risk, practitioners have various sources of information to draw upon. Firstly, they are often provided with details of the offence, sometimes just the category name (e.g., unlawful sexual connection or rape) but most often police summary of facts and/or a judge's sentencing notes as well. Sometimes this is supplemented with notes or reports containing the observations and professional opinions of other practitioners (e.g., previous probation officers, custodial staff, and psychologists) and/or victim impact statements. Secondly, in most cases, the practitioner will also speak to the individual concerned. This typically includes (but is not limited to) first-person reports of their developmental history, the background to their offence, and details of the offending.

Thus, assessments are jointly informed by first-, second-, and third-person perspectives, combined to give a more comprehensive picture of the offence and its putative causes. Finally, practitioners are provided with some sort of template that specifies the structure of the interview, with spaces to fill and boxes to tick, and perhaps most importantly, lists of DRF (in some cases alongside PF) to identify as present (and relevant) or not. In the case of structured risk assessment this is often accompanied by a manual specifying what should be considered evidence that a particular factor is present, details concerning how to reach a final score, and what this score actually means in terms of risk level and targets for intervention. Thus, assessments of various types are informed by multiple sources and perspectives, but the output is most often a list of DRF identified as relevant, perhaps a risk level, and recommendations for action to address these. This prescriptive approach can result in case formulations which look incredibly similar across individuals, as practitioners attempt to fit individuals' experiences within a well-established set of DRF or 'rehabilitative needs' (Ward & Fortune, 2016a).

An important *factual* element of these professional judgments concerns the type of illegal practice concerned and the individuals' responsibility for the crime. This information is provided by the first kind of source, third person descriptions of the crime and judgments of guilt, often by police or the courts. Although the individual may deny their crime (i.e., conflict between sources), it is not the role of the forensic practitioner to assign guilt; the assumption is that the individual has committed a crime. The rest of the judgments are made by the practitioner, based upon other information given (i.e., self-report of the offence), as well as their understanding of the causes of offending. For example, did he intend for this crime to occur? Did he plan it? Why does he think he did it? How does he feel about it now? Does he

think it was wrong? Will he act differently in the future? Perhaps the most important judgment is about the *personal characteristics* or *circumstances* of the individual which *caused* his behaviour, and as such should be corrected in order to stop it from happening again. The practitioner, possibly in collaboration with the individual concerned, makes professional judgments about the presence or absence of each factor based upon their understanding of the DRF, whether it applies to the individual, and if it could plausibly cause the offence. Often the assumption is that 'problems' experienced by large groups of people who have broken the law in general (or a specific type of law) are not only relevant to *this offence*, but also that they have caused or contributed to it.

Figure 1 is intended to illustrate that there are a number of influences upon professional judgments that are largely implicit and rarely acknowledged. Throughout assessments practitioners identify the contexts, subjective psychological states, behaviours, and outcomes which *constitute* the illegal practice. However, this is only part of the picture; there are at least two valuable sources of information missing. Firstly, the wider context contains overarching values (i.e., what is right and wrong, healthy and harmful) and more concrete norms (i.e., what *should* people do or not do) which influence judgments about the nature of offending and related practices. I argue that in order to fully understand offending we must look at the valued outcomes it is directed towards (i.e., what normative tasks is it aimed at?), as well as the collective values/norms it violates (i.e., does it cause harm?). For example, the sexual abuse of a child may be aimed at the experience of connection (i.e., an intimacy related practice), but violates collective values and norms concerning sex and relationships, child and adolescent development, vulnerability, maturity, and capacity to consent.

Secondly, while the identification of DRF points towards broad domains of functioning, it stops short of describing the capacities which they contain and explaining the *mechanisms* which underpin these. To make this clearer, values and norms provide the context within which practices occur (i.e., standards determining their success/appropriateness), and psychological mechanisms cause or enable certain practices within this context. Awareness of the scope of information informing these judgments is a preliminary step in the right direction when it comes to understanding the causes of illegal practices. I argue that this acknowledgement, paired with a deeper understanding of human functioning and the mechanisms underpinning practices, will lead to more comprehensive and individualised case conceptualisations.

In my view, all DRF and PF can be analysed in the same way, and thus all are constituted by practices governed by integrated sets of norms. Values drill all the way down to everyday behaviours and situations. Models of what are viable romantic relationships, healthy

emotional functioning, adaptive attitudes, appropriate sexual interests and behaviour, and acceptable levels of self-control are derived from biological, social and cultural models and ideals. It is to be expected that culturally diverse groups are likely to endorse different relationship practices and normative standards, and that such differences ought to be taken into account when developing risk (and treatment) assessment protocols and constructing programmes.

4.5.3. Case examples: Intimacy practices and self-regulation.

I will now explore two DRF and PF domains; those influencing or constituting intimacy and self-regulation practices, and suggest that sexual functioning, emotion management, and attitudes can be broken down in a similar way. For each domain I will present a case vignette and outline the normative practices that subsequent risk-based judgments rely on, and the values which underpin these judgments. In other words, I will spell out the assumptions concerning what each type of practice *should look like* in prosocial, non-offending individuals, and therefore how deviance is identified and labelled as a DRF in this context. This will allow me to describe each type of practice in terms of its necessary capacities, each spanning a normative range of functioning. My expectation is that the capacities underlying the practices that DRF and PF refer to exist upon these continuums, and that their location is ecologically sensitive and dynamic, rather than being fixed. Each case example (i.e., exemplar) represents an individual who has a primary problem in a particular domain of risk, but it will become apparent that all risk domains (see Table 1) are implicated in both.

Intimacy practices case vignette. Sam has never had a long term emotionally intimate relationship; he has had several short-term dating relationships with women and other purely sexual encounters. When in a dating relationship he has trouble committing to his partner and sharing his thoughts and feelings, eventually driving them away by either ceasing contact or seeking out other women. When asked about past relationships he tends to place the blame on his partners for being "cold", "judgmental", or "manipulative". He states that relationships are a waste of time and that he is happy to play the field, but also reflects that he should settle down at some point (he is almost 35 years old) and find a "good woman" who is worthy of his trust and attention. He would like to have children, stating that they are fun and easy to be around. Sam was recently at a party where he failed to strike up a conversation with any of the adult women present. Feeling rejected, he decided to have a few drinks, and ignore the other adults altogether. He went outside where his friend's 12-year-old daughter was smoking, he decided to join her and while the conversation began in a friendly manner, it ended with him

forcing himself upon her.

Intimacy practices are comprised of one's interpersonal attitudes and skills, preferences for partners (i.e., objects of attraction), the nature of these relationships (i.e., the practices involved), and the contexts they occur within. Intimacy requires a certain level of honesty and trust, acceptance, commitment and companionship, emotional connection, sex and/or physical connection, as well as mutual support and caring (Fletcher, Simpson, Campbell, & Overall, 2013). The aspects of a relationship which can be considered healthy or not include the number and length (stability), and their quality in terms of agreement and conflict. There is general agreement that the object of desire within intimate romantic relationships should be an age-appropriate partner who is able to consent to and participate equally in the relationship. Thus, healthy interpersonal functioning also relies on accurate expectations of and beliefs about different types of relationship. For example, Sam's beliefs about women and his own entitlement are likely to cause problems in the context of romantic relationships with women, with these problems manifesting in short-term relationships characterised by conflict or emotional avoidance.

Mann et al. (2010) identified a lack of emotionally intimate relationships with adults as a potentially psychologically meaningful DRF. A 'lack of' can mean both the absence of enduring relationships and relationships involving "repeated conflict and/or infidelity" (Mann et al., 2010, p. 201). These definitions highlight the fact that 'healthy' intimate relationships between adults have relatively low levels of conflict and are monogamous, however without an understanding of what conflict should look like it is difficult to make a distinction. Similarly, a widely supported DRF for sexual offending is 'emotional congruence with children' (Mann et al., 2010). Emotional congruence involves feeling that relationships with children are more satisfying, that children are easier to relate to or more understanding than adults, and possibly identifying with being a child, for example being emotionally immature. This feature is not evident for all men who have committed sexual offences, but those that do report an emotional congruence with children often speak about their offending as if it occurred within the context of a reciprocal intimate relationship (Mann et al., 2010). For example, Sam's expectation that children are less judgmental than adult women played a part in his decision to approach the girl at the party.

Relatedly, de Vries Robbé et al. (2015) describe the healthy pole of intimate interpersonal functioning as a preference for and capacity to have enduring emotional intimacy with adults. They describe this capacity and preference as "a propensity to form and maintain emotionally close and satisfying relationships with other adults" (de Vries Robbé et al., 2015,

p. 26). Associated PF include: a trustful and forgiving orientation, positive attitudes towards women, honest and respectful attitudes, and care and concern for others (de Vries Robbé et al., 2015). These personal attributes or beliefs appear necessary (or at least beneficial) in the practices of *seeking* prospective romantic partners, *dating*, and *maintaining* a relationship long term. In addition, secure attachment in childhood is considered a developmental PF for both sexual and general offending (de Vries Robbé et al., 2015); theory has long been concerned with the existence of attachment problems and how this might lead to sexual offending for some individuals (e.g., Beech & Mitchell, 2016; Ward, Hudson, & Marshall, 1996). This highlights the developmental aspect of this domain, as we learn about interpersonal functioning and intimacy from our early experiences, both in bonding with caregivers and observing others' relationships. Perhaps Sam's beliefs about intimacy were formed within the context of cold or harsh parenting styles, or earlier relationship experiences which left him feeling rejected and worthless.

Practitioners routinely undertake the difficult task of measuring intimacy preferences and capacities at a single point in time or detecting changes in these after completion of treatment programmes. Judgments concerning whether intimacy related practices are indicative of higher or lower risk of sexual reoffending often rely upon the manifestation of capacities within the context of a long-term stable relationship. For example, risk assessment tools often give a maximum risk score for no current relationship, intermediate scores for co-habiting relationships with problems (significant enough to cause concern to either party) or for a current stable dating relationship, and a score of zero if the individual is currently living with a partner without obvious problems (e.g., Hanson & Harris, 2001). Sometimes these judgments are paired with arbitrary time frames (i.e., living with partner for two years or more), which may or may not be normative for certain generations and cultural groups.

In summary, norms regulating the practices constituting intimate relationships specify what a healthy (and appropriate) relationship should look like; intimate, romantic relationships *should* only occur between consenting adults, *ought* to be reciprocal, *should* include personal disclosure of fears and needs, *ought* to incorporate caregiving and sexual components, *should* contain shared activities and responsibilities, and so on. The terms in italics indicate the normative and value-based elements. Examples of values central to intimacy related practices include; relatedness, connection, pleasure, and the nature of romantic love and relationships (i.e., it is good to be sexually and emotionally intimate with one person for a long time, and without too much conflict). Sam appears to be experiencing problems with both preference and capacity for intimacy with adults. He is fearful and avoidant of emotional intimacy with adult

women (although he is sexually attracted to them), and he lacks the interpersonal attitudes necessary to engage in and commit to this type of relationship. Perhaps this is due to problems with inferring the mental states of women (i.e., they are unknowable or dishonest), or his expectations of intimate relationships based on past learning.

I suggest that the core human capacities which enable engagement in healthy intimacy practices include (but are not limited to): interpersonal skills and self-regulation, attitudes towards the self, others, and relationships (e.g., social roles and expectations), preferences for romantic and sexual relationships, and capacity/desire for emotional intimacy. In this sense intimacy related practices rely heavily on influences from the other four domains of functioning. Impairments in any of these capacities can cause problems in intimate relationships, resulting in high levels of conflict, dishonesty, and/or avoidance of intimacy. In some cases, these impairments and their outcomes (e.g., distress, loneliness, and rejection) can lead an individual to sexually offend against a child in order to meet intimacy needs. However, in other cases they may lead to other more or less healthy coping responses (e.g., dishonesty, promiscuity, substance abuse, self-harm, violence). In other words, many individuals experience problems in seeking and maintaining intimacy, but most do not consider sexual contact with children to be a viable strategy to meet this need. It is necessary to look to other domains of functioning, as well as offending contexts, to construct a comprehensive explanation of the causes of sexual offending.

Self-regulation practices case vignette. Tom is 40 years old, married, and works as a tennis coach with early adolescents. In his mid-20's, and following a sports injury which ended his tennis career, Tom had a problem with gambling and abusing prescription medication, resulting in a large debt and conflict in his marriage. His problems with attaining long-term employment and a tendency to obtain loans for impulsive purchases caused marital conflict. Despite this, things had been generally improving for him until his wife got a new job and began spending long periods of time away from home. Feeling bored and lonely, he began a friendship with a young female whom he coached. She was struggling at home, and he enjoyed feeling like he was helping with advice and support. He gave her his phone number in case she needed anything. One evening after having a fight with his wife, the girl called him upset and he picked her up from her house. They drove to a park where they sat in the car and talked for a while, he kissed her and touched her breasts against her will. She asked to go home and he became angry, accusing her of leading him on. He dropped her off and they continued the conversation via text message. Her mother saw the text messages and contacted Tom's boss. When interviewed about the incident later he reported that he hadn't thought about her like

that before but that he "couldn't stop himself" once he started. He stated that he was upset after the argument with his wife, and that he thought the girl understood and wanted it.

Self-regulation has long been considered relevant to sexual offending; this is reflected in theory and practice targeting (temporary or enduring) deficits in self-regulation abilities (e.g., the *multimodal self-regulation theory of sexual offending*, Stinson, Becker, & McVay, 2016; the *self-regulation model*, Ward & Hudson, 1998b). Self-regulation is "the ability to modulate emotions, thoughts, interactions, and behaviours *effectively*" (emphasis mine, Stinson et al., 2016), and as such encompasses various domains of functioning. Self-regulation is primarily concerned with self-control (i.e., behavioural inhibition), problem-solving, planning, and goal-directed action. It is essentially human *agency*; individuals' ability to intentionally engage with their environment in order to meet their needs. An ability to regulate behaviour (including cognition) is reflected within everyday practices and environments. Individuals' choices, as well as the opportunities afforded to them, largely shape their lifestyles and personal identities. Norms concerning the sort of life people *should* have include central features such as employment and leisure activities, accommodation, relationships, and prosocial participation in society. In this sense, 'good' self-regulation is the ability to live in accordance with the norms and expectations of others, as well as meeting one's own subjective needs. This involves complying with rules and laws (e.g., not having sex with children), as well as norms (e.g., having a job, a home, relationships), and personal values (e.g., academic, sporting, or vocational success). Thus, whether or not an individual is judged to have adequate ability to self-regulate is based upon values spanning multiple levels; ethical, social and cultural, as well as personal preferences. Perhaps the most relevant aspect of self-regulation for offending is the temporary or enduring ability (and desire) to comply with dominant ethical and cultural values and norms based upon the perceived harmfulness of certain practices.

This ability to comply is reflected in the identified correlates of sexual reoffending under the umbrella term 'self-regulation deficits'. DRF relevant to this domain include 'lifestyle impulsiveness' and 'poor cognitive problem-solving' (Mann et al. 2010). Mann et al. (2010) describe lifestyle impulsiveness as low self-control, instability (e.g., employment and accommodation), irresponsible decisions, lack of meaningful daily structure, and problems with long term goals (e.g., limited, unrealistic). These underlying problems can manifest in a parasitic or chaotic lifestyle, including interpersonal conflict, rule/law breaking, substance use, unemployment, and lifestyle instability generally. Unsurprisingly, this DRF is indicative of increased likelihood of reoffending generally and is probably more applicable to some types of sexual offence than others (i.e., some are well thought out and planned, whereas others are

more opportunistic; Smallbone & Cale, 2016). In this example, Tom did not seem to experience lifestyle instability at the time of his offence, as he was employed and had managed to maintain a stable relationship. However, his past problems (e.g., gambling, poor financial management) suggest a propensity to engage in impulsive behaviour when in certain states or contexts.

Poor cognitive problem solving involves impairment in generating and selecting appropriate or effective solutions to life's problems (e.g., interpersonal conflict, financial hardship, unemployment), as well as making everyday decisions which affect the future. Examples of poor problem solving include avoidance, rumination, poor selection (e.g., not considering probable negative consequences), lack of creativity (i.e., a limited selection of solutions to choose from), and an inability to recognise and accurately conceptualise problems as they arise. In addition, an individual may experience difficulties in problem solving when experiencing distressing emotions or while intoxicated, even if their problem-solving skills are generally effective (Ward, Hudson, & Marshall, 1995). This highlights the flexibility and ecological sensitivity of self-regulation ability; problems can reflect enduring deficits or temporary impairments. In the case of Tom above, he seems to have some trouble regulating his behaviour across different situations, as seen in his past employment instability and financial problems. In addition, when in a state of distress, it is likely that it would be more difficult for him to control his behaviour.

De Vries Robbé et al. (2015) list 'goal-directed living' and 'good problem-solving' as two of their proposed protective domains. Another two examples, being engaged in 'employment and/or constructive leisure activities' and 'sobriety', seem to be manifestations of the ability to effectively regulate behaviour in various contexts. Additional factors listed within these domains include self-control, enhanced sense of personal agency, stronger internal locus of control, living circumstances, financial management, life goals, intelligence, and coping (de Vries Robbé et al., 2015). These are in line with the themes above, exercising control or agency, being capable of goal-directed action, and making rational (in terms of being consistent with one's attitudes and goals) decisions – both in daily life, and in demanding or problematic situations. The self-regulation domain covers personal capacities and contexts which support a range of normative practices, as well as the practices themselves which are manifestations of these. For example, having goals, good problem-solving skills, a sense of autonomy, self-control, intelligence, and coping skills facilitate goal-directed living, sobriety, employment, financial management, and so on. Having a balanced, structured, and healthy lifestyle is a manifestation of the capacity for normative goal-directed behaviour (agency) in various contexts (i.e., those that support or offer opportunities to meet goals). Having stable

employment and accommodation, social support, and living a 'good life' (i.e., successful, meaningful) are all dependent on one's capacity to delay gratification, follow rules, and solve any problems which arise in the pursuit of goals.

The second of these PF, good problem solving, is defined as a capacity for managing day-to-day problems as they arise, without becoming emotionally overwhelmed and resorting to unhelpful behaviours. It involves complex cognitive tasks such as combining and evaluating various sources of evidence, considering competing viewpoints, generating numerous possible courses of action, and evaluating the expected consequences of those actions in order to select the best option. It has been suggested that these evaluations are based on three types of expectation; rewards, norms, and competence (Fishbein & Ajzen, 2010). In other words, persons select the most attractive option based upon expected positive outcomes, how other people are likely to perceive behaviour, and their confidence in their ability to be successful. It is likely that this process is supported by the accessibility of healthy coping mechanisms and self-control (i.e., not reacting emotionally), particularly when emotionally distressed. In the case of Tom, the presence of healthy coping strategies or the ability to form accurate expectations about the outcomes of his behaviour could have prevented him from picking up his student.

In summary, the practice of self-regulation is primarily concerned with engaging with one's context and responding to situations in a way that is aligned with one's goals, values, and intentions, as well as the norms and expectations of society overall. There is an expectation that individuals will obey the law and rules more generally (i.e., behaving in a way that is in line with dominant values, not causing harm to others), as well as engaging in normative activities based upon shared values (i.e., having long-term goals, participating in society). Values attainable through these normative practices include mastery of professional and leisure activities, achievement and success, as well as safety and stability in society. People *ought* to be engaged in meaningful employment and leisure activities, to have goals and be disciplined in working towards these, and to live mainly within the boundaries of the law – to control their behaviour and do what is *expected* of them by society. These expectations spell out what it means to be a valuable and productive member of society, and they are reliant upon both personal capacities and the presence of opportunities within one's environment. In other words, persons have varied (internal and external) resources with which to engage in normative and personally meaningful self-regulation practices (Ward & Maruna, 2007).

The internal capacities required for self-regulation include (but again are not limited to) motivation, action-selection, self-control (e.g., inhibiting unhelpful responses), and particular

skills relevant to the goal or value in question (e.g., attitudes, conflict resolution, conscientiousness, etc.). In a sense this domain represents the intersection of the others (see Table 1), as attitudes, emotion management, sexual functioning, and interpersonal skills all come into play within the process of various types of goal-directed practice. These practices can be more or less congruent with one's long-term goals and over-arching values and may be judged by others as appropriate or not. It is important to note that the relationship between self-regulation capacities and sexual offending is complex. Self-regulation capacities can be enduring and relatively stable, but they can also be suspended in certain states (e.g., intoxication, stress), and they can be used to meet needs in various ways. Good self-regulation does not ensure prosocial behaviour and vice versa. At one time or another most people will act in ways that are not in accordance with their own goals and values or those of their cultural group, but usually the outcome is not a sexual offence.

4.7. Conclusions: Causality, Values, and Theory Development

This chapter has provided an overview of the process of theory development and made some suggestions for the enhancement of theories of offending by shifting the focus to goal-directed practices. It began with a causal analysis of DRF which deemed them unsuitable explanatory targets (based upon the issues outlined in the previous two chapters) and outlined a set of capacities and goal-directed practices (see Table 1) which, alongside offending and DRF, should be explained by any comprehensive theory. Thus, this chapter has both argued for the value of mechanistic explanation in adding explanatory depth to theories *and* discussed the reliance of concepts such as crime and DRF on norms and values. Essentially this reflects the difference between dysfunctional psychological mechanisms and the use of functional psychological mechanisms in norm-violating behaviours, both of which are relevant for the explanation of offending. Crime may or may not be caused by dysfunctional mechanisms and contexts which support maladaptive or harmful behaviour. For this reason, it is important to develop integrated multi-level frameworks which can both point to sub-personal mechanisms and explain the role of norms and values in behaviour.

The proposed focus on practices suggests that theory development might be more fruitful if it locates individuals' goal-directed actions within certain normative contexts. As discussed in previous chapters, this is in stark contrast to the traditional focus upon correlates of offending. De Vries Robbé et al. (2015) suggest that "we know very little about what those who have offended sexually value, what makes them happy, and what skills and strengths are related to their desistence from offending" (p. 30). In my opinion, an important potential outcome of a shift in focus towards adaptive practices and their necessary capacities is that it

may prompt further research into the origins and composition of these strengths. It also highlights the fact that the prudential values motivating offending practices are likely to be the same as those of non-offending individuals (Ward & Maruna, 2007).

In summary, whether a feature of an individual or their environment is related to risk is partly a value judgment. These judgments are based upon facts (e.g., the occurrence of violent behaviour, substance use), inferences about that individual's character/capacities (e.g., impulsivity, attitudes, personality), and norms of society overall (e.g., does it fit with the dominant values? Is it healthy? Does it cause harm?). Furthermore, identifying a DRF to target in treatment involves assigning it *causal* status, concluding that it increases an individual's likelihood of offending and needs to be corrected. DRF and PF are defined in relation to the practices that they manifest within – various types of offending and associated behaviours. These practices are goal-directed (i.e., aimed at obtaining prudential values), and evaluated against social/cultural values and norms. Therefore, we cannot explain DRF and PF without reference to the capacities which underpin practices, the goals they are directed towards, and the norms and values which they are evaluated against. We will have a better chance of understanding offending if we look beyond surface level descriptions of 'antisocial' characteristics and contexts, and instead see the whole person, their values, and the normative context within which they strive to grasp these.

The following chapter will develop a preliminary model of human functioning that has utility in explaining crime, based upon agency and its associated abilities. The hope is that in conceptualising DRF and PF as aspects of agency, we can shift the focus to the *individual* rather than criminal behaviour and its correlates. In discussing agency, I refer to an individual's capacity for, and engagement in, intentional, goal-directed action. The capacities underpinning agency contain various mechanisms (i.e., entities and processes) spanning multiple levels of explanation, enable agency processes, and underpin the phenomena associated with offending (i.e., DRF and PF).

Chapter Five: Prediction and Agency

I have so far outlined the problems with current explanations of offending which rely upon the concept of dynamic risk factors (DRF) and suggested that explanation would be better directed towards the kinds of practices within which these problems manifest. This will go some way to addressing the normative nature of DRF and allow for cultural and individual differences in the expression of these problems. The aim of this chapter is to further address two questions: 1) what are DRF and PF? And 2) what possible capacities underpin these phenomena and are thus associated with criminal behaviour? This chapter will build upon the focus on goal-directed behaviour, through developing a model of agency capacities and processes. This model conceptually sits a level below theories such as the *general personality and cognitive social learning approac*h (GPCSL; Andrews & Bonta, 2010) and typologies of protective factors (PF; de Vries Robbé et al., 2015; Thornton et al., 2017); it adds depth by depicting the process of behaviour and pointing to mechanisms for further explanation. This preliminary model integrates theory across disciplines, spanning multiple levels of analysis, and hopefully offers a useful (partial) explanation of criminal behaviour and a better understanding of how DRF and PF are related to offending.

In order to begin to address the conceptual issues with PF (see chapter three), Ward (2017) developed the *predictive agency model of protective factors* (PAMP). This model combines ideas from evolutionary biology, metaphysics, cognitive neuroscience, and psychology to form a tentative, comprehensive account of PF, which equally applies to DRF. This conceptual framework provides the basis for the model presented here, which will be renamed simply as the *predictive agency model* (PAM; Heffernan & Ward, 2017) to reflect the fact that it applies equally to all types of goal-directed behaviour, regardless of the outcome. It is important to clarify that not all goal-directed behaviour is intentional or conscious, and habitual, routine, or even basic cognitive processes such as perception may involve goals without being subject to awareness. My hope is that DRF, PF, and offending can be explained as variation, and in some cases dysfunction, within the abilities required to meet human interests.

5.1. A Predictive Agency Model

The PAM is a theoretical framework that aims to provide an explanation of PF and DRF and their associated phenomena, based upon Ward's (2017) earlier work. Thus, it shares the theoretical commitments and core components of the PAMP but is applied to goal-directed behaviour in general, and criminal behaviour (including all features and arenas classified as

risky or protective) in particular. I will first outline the theoretical commitments which underpin the PAM, and which should constrain future work to develop the model. These are broadly drawn from the above disciplines, but in particular rely upon the concept of *prospection* outlined in a recent exploration of human functioning and goal-directed behaviour (Seligman, Railton, Baumeister, & Sripada, 2016). I will then present an example of how the PAM can explain the occurrence of DRF alongside offending, and briefly suggest that it can also account for the development of PF and desistance processes.

5.1.1. Theoretical commitments of the PAM.

The first major theoretical commitment of the PAM is to the evolutionary origins of *subjectivity* in the form of first-person perspective (see Ward, 2017). The development of subjectivity in animals most likely evolved over the course of several million years and was selected for because of its ability to promote survival and reproduction goals (Walsh, 2015). The ability to find the way back to food caches, track prey, avoid predators, and to seek mates requires an animal to adopt a rudimentary first-person perspective and know where it is at any particular time relative to other animals and places (Neisser, 2015). This ability rests upon possessing relatively accurate or adaptive 'general models' of the self, others, and the world; which represent reality, including opportunities and constraints (i.e., whether the environment supports or thwarts the attainment of goals). Human beings' orientation in place and time, the capacity to plan and predict other peoples' behaviour, and to think conditionally with respect to the future arguably originated from navigational skills and affective engagement in the world (Barrett & Bar, 2011; Neisser, 2015; Seligman, et al., 2016; Walsh, 2015).

A second major commitment of the PAM is to the critical role of *emotional systems* in the development and operation of agency. Emotional systems help organisms identify potential resources such as food and other types of beneficial objects and events; emotions track valued outcomes and states. Without some way of affectively tagging situations or possible outcomes it would be difficult to engage in effective decision making and action at all. In brief, emotions are organism wide responses, initiated without intent, involve evaluations, are motivating (i.e., ready the organism for action), are generally adaptive, and have meaning; they are in effect, *sense making* (Maiese, 2011). Emotions frame salient aspects of a problem task and help to focus attention and subsequent problem-solving efforts on these features. Thus, emotions have their evolutionary and developmental origins in the enabling of agency and search for meaning. They are inherently evaluative and reflect what people care most about.

The third major theoretical commitment is to a multilevel, *nested systems* view of human beings, and the implications of this view for the explanation of crime and desistance. Once it is accepted that the nature of human beings is formed from the dynamic interplay between biology, psychology, culture, and social experience it is obvious that traditional criminology can only offer a partial understanding of normative breakdowns such as crime (Durrant & Ward, 2015). As *embodied and evolved organisms*, individuals' actions can only be explained by the consideration of biological as well as social and psychological causes (Durrant & Ward, 2015; Maiese, 2016). The molecular and neurological levels of analysis are as important as the social and psychological. Different disciplines offer unique perspectives on human functioning which can be combined (via integrative pluralism) into *multilevel* and *interfield* explanations of psychological and social phenomena such as crime. Although, different levels of analysis and their associated mechanisms may be more salient in particular circumstances. For example, the relationship between drug addiction and crime may be best explained at a molecular and/or neural level. While the PAM is aimed primarily at the first person and psychological levels of explanation and does not explicitly outline each level of analysis, it will be important for future work to explore the biological and neurological underpinnings of the psychological phenomena outlined in the model, as well as the socio-cultural level.

A fourth, related theoretical commitment is that subjective or first-person level explanations are *irreducible* when attempting explain individuals' actions, including crime. They are irreducible because as argued above, the systems constituting human beings are nested within each other and collectively comprise a human person. Each level contains structures and processes that interact to produce novel properties that do not exist at another level. For example, an intact neuron requires a cell body, dendrites, axon, myelin sheath, appropriate concentrations of calcium and sodium, and so on; collectively these components enable the neuron to function as a unit. If neurons are organised into networks and communicate, additional properties emerge which are not possessed by their components, for example, spatial feature detection. Networks of neurons clustered into even larger functional systems confer additional abilities such as face recognition, autobiographical memory, and motor control. Furthermore, the psychological and biological systems constituting a person are able to achieve things previously impossible once they are embedded in larger systems with norms and social rules (e.g., playing games or buying a house). New properties emerge from the mechanisms operating at each level, creating new abilities and novel properties. As persons acquire language and an understanding of social norms, they develop a self-concept, self-

narratives, and reflective abilities. Personal or intentional level explanations are unable to be eliminated because persons are real entities that exist in the world (Baker, 2016; Maiese, 2016; Zhavi, 2006). These explanations provide valuable information about the likely immediate and future actions of individuals and make it possible for groups to form collective plans, and ultimately, social institutions (Sterelny, 2012; Suddendorf, 2013). In practice, this is the level at which we engage with each other. Practitioners gather information from the viewpoint of persons who have committed crimes and use this information to make decisions concerning risk and form a therapeutic alliance.

The fifth theoretical commitment of the PAM is to a set of related capacities and their internal and external components that enable the mind to function as a *predictive engine* (Bar, 2011; Clark, 2016; Hohwy, 2013; Seligman et al., 2016). The first-person perspective provides human beings with a unique set of predictive capacities involved in motivating action, attuning persons to threats and benefits (i.e., emotion), setting priorities, values, and expectations, and enabling reflection and modification. The key idea is that the mind/brain evolved to efficiently utilise internal and external cues and knowledge (i.e., general models) to predict outcomes and thereby optimise the chances of achieving survival related goals (Clark, 2016; Hohwy, 2013). Basically, the application of general models to explain and predict outcomes enabled people to exert greater levels of control over their own actions, other peoples' actions, and events in the world (i.e., to make and implement plans). Researchers have used a variety of terms to refer to this ability such as 'prospection', 'simulation', 'prediction', and 'expectation', but the meaning is basically the same. Seligman et al. (2016) expressed the core meanings well in their definition of prospection "as a label for the mental process of projecting and evaluating future possibilities and then using these projections for the guidance of thought and action" (p. 6). The mind/brain 'fills in' the missing bits of information by making guesses based on prior learning, embodied in internal, dynamically evolving general models. This is an efficient use of information, as regularities in the world mean that we can make informed *guesses* without processing every piece of data. For relatively vulnerable organisms such as homo sapiens this was a significant innovation (Sterelny, 2012; Walsh, 2015).

Human beings' selective advantage resides in their ability to model the environment and make informed predictions about the future responses of the self and others in any given situation. According to the predictive mind/brain approach, human beings evolved to be dynamic learners whose facility to adapt depends on their ability to construct general models that accurately predict the existence of objects, other peoples' mental states and behaviour, and future events. In addition, the capacity to construct, test, and revise general models enables

people to evaluate themselves and the outcomes of their actions. Individuals acquire these capacities and models through personal experience and social/cultural learning; thus persons are partly social in nature. They are essentially forward-looking animals who are by nature predisposed to construct plans and predictions to guide behaviour. Adaptive functioning requires the acquisition of internal and external capacities and resources to achieve valued outcomes and to avoid and/or escape from potential harm.

5.1.2. The PAM.

The five theoretical commitments outlined above cohere with current scientific understanding of psychological structures and processes and their relationships to social action. The PAM (see Figure 2) is constructed out of the above theoretical commitments and can be applied to DRF and PF in the criminal justice domain.
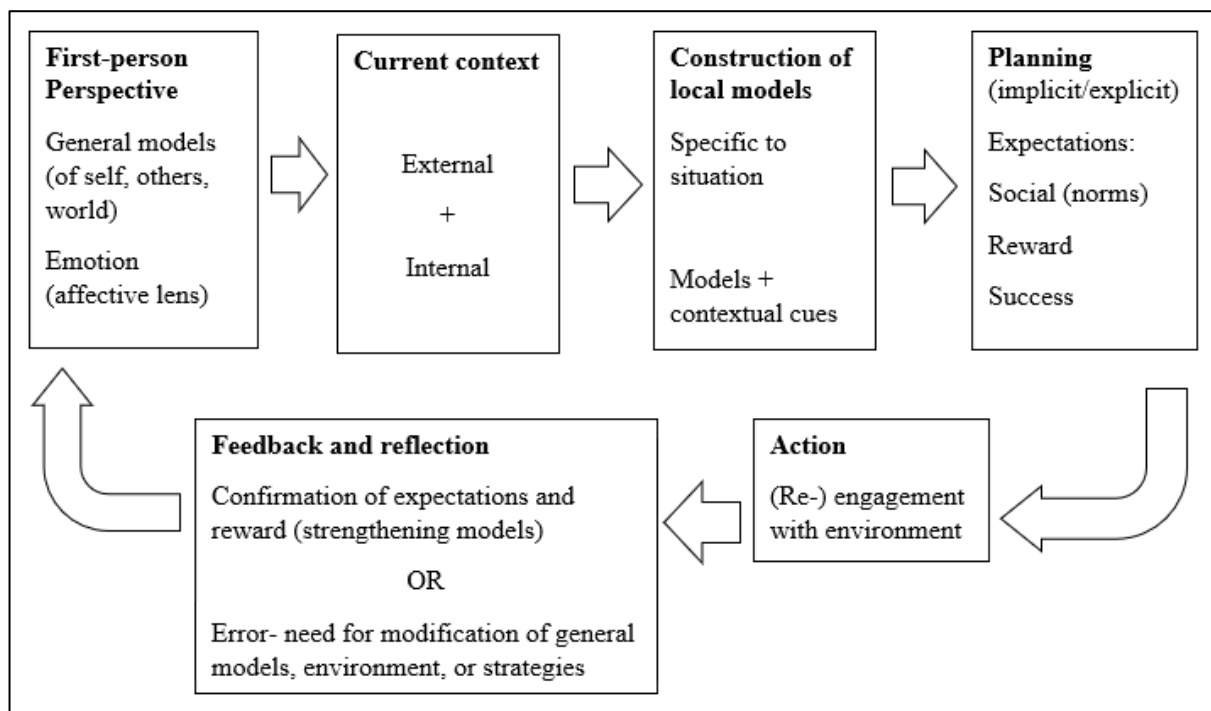
Figure 2. Predictive agency model.[4]

Subjectivity or a *first-person perspective* is comprised of both general models and emotion. This capacity allows human beings to create convincing simulations (i.e., of future situations and possible courses of action linked with affect), which are crucial for long term thinking. The agent needs to both project the self into the future (i.e., imagine possible actions

---

[4] Reprinted from "A Comprehensive Theory of Dynamic Risk and Protective Factors," by R. Heffernan and T. Ward, 2017, *Aggression and Violent Behavior, 37*, p. 137. Copyright 2017 by Elsevier Ltd. Reprinted with permission.

and outcomes) and simultaneously link these simulated possibilities with authentic emotion (e.g., one needs to *feel* what that future situation will be like). Seligman, Railton, Baumeister, and Sripada (2013) refer to these as 'if-then conditionals', they are general models drawn from past experiences in which the individual has learned what to expect from various situations and actions. Particular contexts or situations trigger the retrieval of relevant general models, which are used to construct situation specific local models. The agent makes sense of their current experience by mentally representing the situation, others, opportunities and constraints (based on general models and emotion). It is suggested that constraints on cognitive resources can affect the ability to generate local models, as well as the planning and subsequent phases (Seligman et al., 2013), possibly explaining why some people only commit offences when experiencing distress or intoxication.

The planning phase involves (implicitly or explicitly) generating possibilities for action and representing likely outcomes, and emotion helps to evaluate these options (i.e., they are affectively tagged based on past experience). It is likely that many options will involve both positive and negative outcomes (to some degree), and that these will be more or less in line with an individual's values and priorities (i.e., emotions track value). Generating multiple possible options for action is an important human capacity (i.e., we experience autonomy or 'free will' in most situations). Thus, it is difficult to pinpoint a cause of behaviour, because it rests on the generating and evaluating of options, which is dependent upon a range of capacities. The selection of behavioural options is reliant upon both internal and external resources (i.e., they draw from past experience and present opportunities to construct possible futures), and not everyone possesses the same resources for agency. After an action has been selected and completed, feedback and reflection help the individual to experience the consequences and make sense of these. This process relies upon values, emotions, and flexibility of general models. For example, sometimes an individual may experience negative internal or external feedback (e.g., detection and arrest and/or guilt and shame) and alter general models, or other times the costs of chosen options can be compensated for through justifications or minimisations, avoiding the need to change general models (i.e., view of the self or others).

Within the PAM, DRF and PF are best thought of as manifestations of variation (i.e., strength or weakness) in any of the above capacities (including internal and external components) and processes. These include interpersonal and intrapersonal capacities (e.g., empathy, subjectivity), emotion, general and local models, prediction, and self-regulation. In order to clarify this rather abstract model I will now work through an example (see Figure 3). An individual's experience of the DRF category of 'intimacy deficits' (or relationship issues,

general social rejection, social isolation, etc.) and motivation to achieve intimacy can be explained by appealing to each aspect of the model.
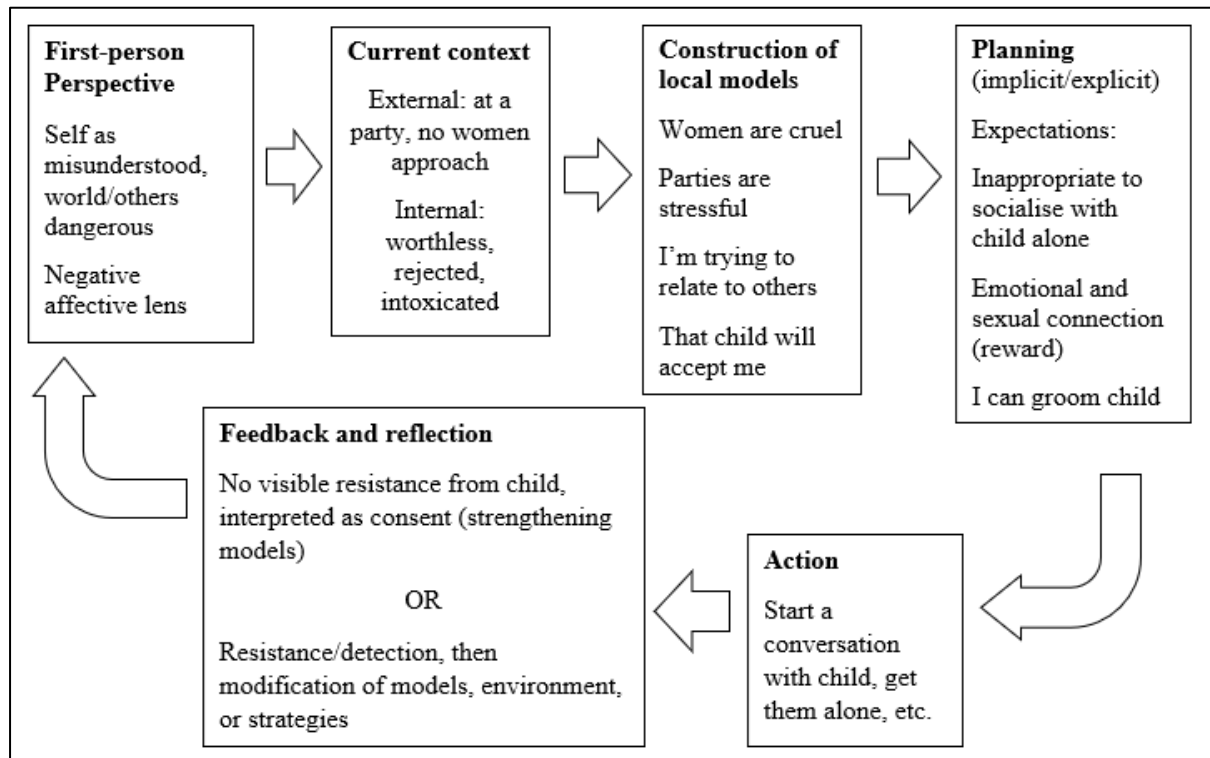
Figure 3. Predictive agency model with example.[5]

Firstly, *general models* of the self as vulnerable or misunderstood, and female adults as dangerous and deceptive are triggered or activated in contexts where the individual seeks intimacy but experiences rejection. In addition, an *affective lens* that is generally negative is more sensitive to cues that make the individual feel rejected, unloved, and lonely. Within this context (e.g., a social gathering) *local models* are implicitly constructed to represent the specific occasion and people present, and the behaviours which are acceptable. Past attempts to experience intimacy have repeatedly failed and resulted in frustration, so over time he has come to consider alternative goals and plans (e.g., intimacy with anyone). His plans are based on his subjective *expectations* of his social environment (i.e., what others will see/think of his behaviour), the rewards available (i.e., whether the outcome will be satisfying), and whether or not he believes he possesses the skills required to successfully enact his plan. It is important to note that the planning process can be explicit and well thought out, or implicit and seemingly

[5] Reprinted from "A Comprehensive Theory of Dynamic Risk and Protective Factors," by R. Heffernan and T. Ward, 2017, *Aggression and Violent Behavior, 37*, p. 137. Copyright 2017 by Elsevier Ltd. Reprinted with permission.

automatic. He then puts his plan into *action* in order to test these expectations and experiences the consequences of this behaviour. The *feedback and reflection* process can either involve experiencing the desired outcome (reward) and strengthening the general and local models or it can signal error. When outcomes are experienced as rewarding, plans are seen as successful and the likelihood that the behaviour will be repeated increases.

Alternatively, in situations where something has gone wrong (e.g., inability to achieve desired outcome, victim resistance, or detection by other adults), the individual has several options to enhance his predictive agency capacities. He may over time modify his general models, which in turn alters his ability to predict situation-specific outcomes (i.e., generate local models), form expectations, and make plans. Otherwise he may attempt to manipulate his environment to fit with his existing general models or reformulate strategies so as to avoid negative consequences. It is important to note that enhanced predictive agency does not necessarily lead to reduction in offending, more accurate general models and related interpersonal abilities can be used to gain expertise in offending. However, in cases where feedback and reflection lead to more adaptive, prosocial, legal, and less harmful plans and behaviours, we might say that PF are present. These PF can be developed over time, for example through treatment or as an effect of maturation and changes in circumstances. I will now briefly outline one way that the PAM can account for PF reducing the likelihood of offending.

Firstly, a capacity for intimacy with adults may develop with a therapeutically induced shift in *general models* about the self and others. For example, seeing the self as worthy of love and affection, other adults as capable of providing this, and children as unable to participate in emotional and sexual relationships. It is also possible that emotional processes may change, for example through medication or therapy. If viewed with a generally positive *affective lens* and high levels of self-esteem, even contexts featuring social rejection are less likely to be experienced as personally defining (i.e., they do not fit with the subjective self-concept) and so are less likely to trigger feelings of worthlessness. The construction of *local models* in these environments may then represent *specific individuals* as particularly hostile, or as romantically incompatible with the self. The individual's plans for engaging with the social situation may then involve either altering his conversational strategies, approaching other people, or leaving the situation in favour of a more accepting group of peers. His capacity for intimacy with adults (rather than seeing them as dangerous) has led to expectations and behaviours that are consistent with this, even if he does not experience a connection on his first attempt. This predictive agency process occurs constantly, hence action can be thought of as re-engagement

with environments and situations that the agent has experienced before, and general models, expectations, and other capacities shift or develop over time – directing the individual towards or away from crime.

5.2. Preliminary Evaluation of the Predictive Agency Model

I will now outline the strengths and weaknesses of the PAM relative to other theories of DRF and offending, and comment on its ability to address the conceptual issues outlined earlier. The aim of this chapter was to address two preliminary questions concerning the nature of criminal risk: 1) what are DRF and PF? And 2) what possible capacities underpin these phenomena and are thus associated with criminal behaviour?

In answering the first question, according to the PAM, DRF and PF are broad categories containing (proxies for) weakness or strength in the capacities underlying predictive agency (e.g., emotion, general models, prediction) that cause behaviour that is more or less harmful and/or illegal in particular contexts. They are context-bound; what may be a strength or weakness for one individual in one context may not be for another, and they do not relate to reoffending or desistance in any straightforward way (i.e., strengths are not necessarily prosocial and vice versa). The problem of broad definitions and composite constructs is addressed here by locating DRF and PF within a theory of human agency, which when paired with knowledge about the perspectives and experiences of individuals, has the potential to identify specific areas of concern (i.e., weaknesses to target). I propose that it is not necessary (or perhaps even possible) to come up with a set of specific causes for various types of crime. Instead offending is better understood within the context of human agency, and people's orientation towards (expected) rewarding behaviours. The PAM focuses on the specific goals individuals have and how they go about securing them in particular contexts. This means that the explanatory attention is directed towards important domains of functioning underpinning tasks or needs such as intimacy, mastery, control, and so on. Features typically referred to as PF assist people to achieve these goals in legally acceptable ways while DRF (i.e., impairments in any of the above components of agency) make this more difficult.

In comparison to recent theories of DRF and PF discussed in chapter three, the PAM arguably adds value by outlining the processes by which action is motivated, and subsequent decisions are made based on expectations (predictions) of specific situations and actions. In this sense it has greater explanatory depth, as it adds the concepts of subjectivity (first person perspective) and general and local models, and points to the role of predictive processes and expectations across offence-related action-sequences. The context also features heavily in the

PAM in prompting the formation of local models, thus acknowledging the ecological sensitivity of human beings. Thus, it is compatible with the agency model of risk (Heffernan & Ward, 2015) and the agency filter model (Serin et al., 2016), but it adds more detail to the action selection process and outlines the role of emotion in giving meaning to various experiences.

In addition, while being consistent with Thornton's (2016) theory of DRF in drawing upon motivation (in the form of emotion) and reasoning (i.e., the *theory of reasoned action*; Fishbein & Ajzen, 2010), it goes further by offering context-specific local theories as an alternative to his overly general schema. Thornton (2016) offers a new way to think of DRF as problems in goal-directed reasoning (going deeper than the propensities model), but there is still room for "researchers to explore the relationship between risk-related propensities and more fundamental factors on which the propensity depends" (Thornton, 2016, p. 147). In applying his theory to PF and creating four categories to outline the types of factor that this concept encompasses, Thornton et al. (2017) demonstrated that PF largely consist of arenas where protective processes may take place. However, it seems to me that the PAM can better depict dynamic internal factors interacting with (social and professional) contexts across the desistance process. For example, how dynamic internal PF influence responses to environmental cues, expectations, and the evaluation of behaviour across time. Finally, affective processes are relatively neglected in Thornton's (2016) model whereas they play a crucial role in the PAM in orientating persons towards or away from (expected) valued outcomes.

Durrant's (2017) evolutionary account of PF is useful in understanding distal causes (i.e., the onset of offending or abstaining from crime), but does not say much about the current state of the individual (i.e., proximal causes of reoffending). The PAM is able to highlight the proximal causes of offending and explain changes in patterns of behaviour over time (i.e., the feedback and evaluation process). In this sense it is more useful in guiding practice, as it can indicate areas of vulnerability and potential strength. While it is helpful both theoretically and from a prevention perspective to know that life history strategies and plasticity can explain individual differences, it is difficult to see how this may inform practice. Thus, the PAM adds to our understanding of current antisocial behaviour and is also compatible with this framework in the sense that it acknowledges the evolutionary origins of subjectivity and prospection.

Furthermore, the PAM is compatible with the range of influential single and multifactorial theories (see chapter three) aimed at explaining offending, while offering greater explanatory breadth. That is to say that it can explain a wider range of phenomena than is

typically depicted within narrower theories of offending. It also adds explanatory depth, for example, one major advantage of the PAM is the critical role of the emotional lens (i.e., in giving meaning and guiding action). Emotion affects the way we view the world, how we experience different situations, and what sort of outcomes we view as rewarding. The meanings attached to certain situations or people may be considered adaptive or maladaptive, depending on their outcomes in various contexts. To offer a different example, if property crimes are experienced as a source of mastery experiences, someone might seek out situations where they are able to refine and exercise their theft related skills (e.g., sharing stories with others, casing houses, studying security systems online; Nee & Meenaghan, 2006). If a person has been introduced to theft at a young age, the association between property crime and feelings of mastery and achievement may motivate them to increase their rate of offending and difficulty of targets to demonstrate their superior expertise (partly an affective phenomenon). More specifically, the anticipation of committing a crime (mounting sense of excitement), the thrill of approaching a house (pleasure, joy), the exhilaration of 'cracking' a complex security system (elation), and the feelings of pride after successfully robbing a house play crucial roles in the planning and execution of the offence. By tracking emotion we are better able to understand the decision-making processes and preferences that shift individuals either towards or away from harmful and criminal behaviour – the process gains meaning.

Since its development the PAM has demonstrated fertility or heuristic value. For example, it has been built upon to explain the role of culture in behaviour and to explore the reasons for cross-cultural differences in DRF (Schmidt, Heffernan, & Ward, 2020; Strauss-Hughes, Heffernan, & Ward, 2019;). The *cultural-ecological predictive agency model* (CEPAM; Strauss-Hughes, Heffernan, & Ward, 2019) highlights the influence of culture as shaping persons and contexts. It also acknowledges the intersectionality of culture as being shaped by different aspects of persons, their environments and their identities, rather than equating culture with ethnicity. The CEPAM offers two additional layers of influence on the agent; *historical contexts and processes* and *cultural systems*. The model suggests that persons are embedded within historical contexts which have shaped their current sociocultural and physical context. For example, New Zealand's history of colonisation has shaped the current cultural context, including its institutions, laws, and attitudes of society. These have resulted in adverse outcomes for indigenous persons, such as marginalisation and over-representation in various health and crime statistics. These experiences (both historical and contemporary) have influenced the way that individuals belonging to various groups within society view themselves, others, sources of reward, and so on. In other words, cultural processes and

contexts both shape the content of general models and their use within goal-directed processes (e.g., perception, interpretation, planning, reflection), as well as shaping the context within which the agent acts (e.g., opportunities and constraints in current contexts, others' perceptions, feedback). In this sense the PAM has the potential to explain variation in offending and DRF across cultural contexts, it is widely applicable and can account for individual differences, some of which relate to group membership.

I would like to, at this point, discuss a general criticism of theories of DRF and offending, which may also be applied to the PAM, somewhat due to its broad focus on behaviour. The issue is that theories are largely unable to point to the reasons why the specific behaviour is chosen as the means to meet the need (i.e., the planning stage of the PAM). Theories of offending often miss the mark because of their focus on DRF, which are not specific to offending (i.e., many people experience anger, frustration, intimacy problems but do not engage in sexual offending or violence). The PAM intentionally directs explanation at goal-directed behaviour in order to sidestep this problem with the reliance on DRF. However, in doing so, it may be suggested that it also cannot explain why someone chooses to sexually abuse a child to achieve inner peace or cope with distress (for example), instead of using substances, hitting their partner, engaging in self-injurious behaviour, setting fires, etc. I suggest that this complex issue can be partially addressed by the PAM.

According to the PAM, this variation in problem behaviour (both between individuals and within the same individual across varying contexts) is essentially a result of behavioural options generated during the planning stage. These options are informed by general models which have developed over time through experiential learning and observation. This learning is a product of both *exposure* to different opportunities, constraints, and situations, as well as the *meaning* attached to these by the individual (i.e., via emotional processing). Human beings are not passive recipients of information, rather we engage with our environment through a process of cultural learning whereby we make sense of and attach meaning to our experiences (i.e., forming general models of ourselves, others, and the world). This means that individuals are affected differently by events, in line with the agency models discussed earlier (e.g., evolutionary approaches and the agency-filter model; Durrant, 2016; Serin et al., 2016). Therefore, an overly simplified answer to the question of why some individuals choose to sexually offend is that this behaviour was an available and attractive option for them at the time. This may be due to long-standing general models such as 'children are sexual beings' (e.g., learned through past experiences of abuse or exposure to child exploitation materials) or temporary states where this is an option when it usually would not be considered (e.g., through

the faulty construction of local models due to stress or intoxication). The way these models develop over time or arise seemingly spontaneously should be a key focus for future research.

Despite its merits, the PAM is very general, and the use of additional theory, knowledge, and reasoning are required to make sense of individual cases. Further theoretical work is required to advance and add depth to the model. In particular, the biological aspects of the model are under specified. The second question posed in this chapter (concerning potential capacities or mechanisms of agency) remains open and requires further research. This research should follow the guidelines for theory construction outlined above (and discussed in the following chapter), and will add depth to the capacities required for agency in the PAM. A preliminary list of these capacities, their internal and external elements, and examples are contained within Table 2 below.

Table 2

Agency Capacities and Components[6]

Capacity	Internal components	External components	Example DRF/PF
Emotion	Valence (affective lens) Internal cues/signals Meaning Motivation Emotion recognition and management	Environmental triggers or signals (e.g., threat, reward)	Negative affect vs. trustful and forgiving orientation
General models	General mental representations of self, others, the world e.g., self-narratives, identity Local models	Social learning, past experiences supporting or challenging general models, norms	Offence-supportive attitudes and beliefs vs. attitudes supportive of respectful and age appropriate relationships Antisocial and offence-supportive identity vs. prosocial identity
Planning	Learning and memory Attention and perception Planning, i.e., cost vs. benefit Flexibility/creativity	Social and physical environments, opportunities, norms, learned expectations	Poor problem- solving vs. goal-directed living
Action	Impulse control Thought control Language Communication Theory of mind	External barriers Access to social networks and relationships	Self-regulation Impulsivity vs. self-control
Feedback and reflection	Learning and memory Reflectiveness	External feedback	Lack of empathy vs. constructive social and professional support/control

These abilities are consistent with our understanding of offending behaviour specifically and human beings in general. The specific (internal and external) resources and examples chosen for inclusion in Table 2 are intended to illustrate possible components or

---

[6] *Note.* Adapted from "A Comprehensive Theory of Dynamic Risk and Protective Factors," by R. Heffernan and T. Ward, 2017, *Aggression and Violent Behavior, 37*, p. 139. Copyright 2017 by Elsevier Ltd. Adapted with permission.

mechanisms within each capacity, they are by no means fixed or exhaustive. This provisional picture of agency capacities needs substantial investigation in order to flesh out local theories within these domains and link them within a model of goal-directed behaviour (i.e., the PAM or similar). These local theories should be informed by a range of psychological disciplines and contain explanations at various levels of analysis. For example, each of these abilities should ideally contain biological, neurological, phenomenological, experiential, and contextual elements.

5.3. Conclusions: Predictive Agency Model

In this chapter I presented a preliminary model which is consistent with existing theory and able to account for different combinations of DRF and PF, and associated behaviour. According to the PAM, these factors are more or less problematic features of the processes and capacities constituting predictive agency. In other words, they contain or are underpinned by capacities or resources that exist on continuums and make it more or less likely that an individual will use harmful behaviour to achieve their goals. Further research should focus on the composition of agency capacities. For example, clarifying the nature of the emotional processes that guide predictive agency. A comprehensive explanation of the formation, activation, and modification of general models will also add value to the model. Well-researched phenomena such as learning, memory, decision-making, emotion, motivation, communication, and theory of mind (among many others) will have much to offer in developing these deeper explanations.

Chapter Six: The Risk-Causality Method

Building upon the *predictive agency model* (PAM; Heffernan & Ward, 2017) and other agentic approaches to explaining behaviour, this chapter will put forward a methodological framework to add depth to these theories. It has been suggested recently that DRF in their current form are best thought of as (at least partially) symptom-like features of individuals and their environments which are generated by causal mechanisms rather than being causes themselves – they are a good starting point for reasoning about the causes of offending (Ward & Beech, 2015; Ward & Fortune, 2016a, 2016b). A difficulty with incorporating DRF into explanations of (re)offending, risk assessment, and/or forensic case formulations is that it is assumed that they are coherent constructs (e.g., Hart, Sturmey, Logan, & McMurran, 2011; Sturmey & McMurran, 2011). In my view this is not the case, and DRF are better conceptualised as red flags (i.e., symptom-like features) that indicate the existence of problems but cannot explain why they have arisen or persisted. They may be useful predictors but are not coherent explanatory constructs. The current emphasis on empirically established lists of DRF or criminogenic needs embedded within a general personality and cognitive social learning theory (Andrews & Bonta, 2010; Bonta & Andrews, 2017) in both research and practice falls short of causal explanation (see Ward & Fortune, 2016a). Further, while theories of crime sometimes incorporate specific components of DRF, in my view, this is not as systematic it could be.

The major goal of this chapter is to explore ways in which DRF may be useful in developing explanations of offending. I offer one possible method for utilising DRF within explanations in research and eventually practice, the *risk-causality method* (RCM; Heffernan, Ward, Vandevelde, & Van Damme, 2019). I will begin by discussing a number of possible solutions from the recent theoretical literature, and comment on their potential and limitations in addressing the problems currently faced by DRF. Then I will present the RCM in detail and apply it to the example of 'emotional congruence with children', a sub-feature of the intimacy or interpersonal domain of DRF. I will conclude with some remarks concerning its potential contribution to the field. It is hoped that the use of the RCM can make the task of translating DRF into specific causal processes easier, and therefore bridge the current theoretical gap between prediction and explanation in both research and practice domains. As previously noted, I primarily refer to the area of sexual offending to illustrate the argument. This is purely a matter of convenience and the problems noted in the ways DRF are conceptualised and used in research and practice generalise to other types of offending.

6.1. Incorporating Dynamic Risk Factors into Forensic Explanation

There are a number of suggestions put forward for how to best utilise DRF in the explanation of (re)offending and treatment planning. I will consider five of these here: 1) ignore the above problems and treat *DRF as causes*; 2) evaluate them against a set of *risk factor causal criteria*; 3) utilise a *risk-matrix*; 4) reconceptualise DRF as (proxies of) impairments in agency; and 5) locate DRF within social *exemplars/practices*. The conclusion is that there is potential value in combining these ideas within a broader methodological framework (the RCM). Most of the frameworks and ideas discussed in this section have appeared only relatively recently in academic journals and books. In part this is because the conceptual status of DRF has not previously been questioned and it was simply assumed that some of them (at least) were causal constructs.

6.1.1. DRF as causes

The first strategy involves accepting DRF in their current form and relying on these as promising candidates for the explanation of (re)offending – essentially business as usual. However, due to the difficulties outlined above I do not think that this is a viable theoretical strategy, and I suggest that it will eventually lead the field of forensic and correctional psychology into a theoretical dead end (Ward & Beech, 2015). This stance is supported by the previous chapters of this thesis outlining the empirical and conceptual issues which preclude DRF from being accepted as causes of offending. Some of these issues are listed in the following section on causal criteria.

6.1.2. From DRF to causal status: Bradford Hill's criteria

One strategy which rejects the assumption that DRF are causes in their typical composite form, asks researchers to evaluate the potential causal elements of DRF against standards such as Bradford-Hill's (1965) criteria for causal inference in the medical epidemiology field. These criteria were outlined at the beginning of chapter four, and I concluded that DRF in their current form arguably satisfy a number of these criteria, however they fail to meet others. To summarise, in their support, they are empirically derived risk correlates and have been observed to precede reoffending in longitudinal studies (Bonta & Andrews, 2017). However, they do not appear to hold across offending groups and raters; they are not specific to offending populations, they do not *reliably* exhibit the expected increases and decreases in relation to recidivism rates, they are not theoretically coherent and do not refer to causal mechanisms, and analogous concepts have not demonstrated causal relationships. The problem is that unlike the questions of causality posed by epidemiology (e.g., the relationship

between the risk factor of smoking and lung cancer), risk factors do not refer to causes in any direct sense and are inherently vague – in their current composite form they could refer to any number of possible situations and properties. Without further conceptual and theoretical analysis they simply contain too many causal possibilities to be confident of their role in facilitating norm-violations. In addition, many of the elements of causality contained within DRF categories have not yet been researched extensively with offending populations, and so it will be difficult to ascertain whether they meet the above criteria.

The criteria outlined above offer useful suggestions for ascertaining whether or not specific factors summarised by DRF can be *justifiably considered* by researchers and practitioners as possible causes or not. However, because these guidelines do not directly assist in the identification of the mechanisms underlying DRF, in the criminal justice area they are best utilised as initial filters to help identify possible causes of crime and its problems. For example, if they are reliably associated with and precede offending (at a minimum) then they might point to a potential mechanism generating problems in prosocial goal attainment. However, we still need to uncover this potential mechanism and explain its relationship with behaviour.

6.1.3. Dynamic risk research framework: Risk matrix.

A third suggestion for the investigation of DRF was put forward by Ward and Fortune (2016b), the *dynamic risk research framework* (DRRF). This approach involves identifying the causal elements of DRF by referring to the psychological processes inferred in recent psychopathology research (i.e., it is heavily based on the research domain criteria or RDoC project matrix; Lilienfeld, 2014). Briefly, the aim of the RDoC project is to develop new ways of classifying mental disorders based initially on five domains of psychological processes and their instantiation in neurobiology. The DRRF proposes that potential causal processes should be teased apart within a matrix spanning multiple levels of analysis. The Y axis of the matrix contains six categories of possible causal processes: *negative affective systems, positive affective systems, cognitive systems, self-regulation systems, intrapersonal social systems,* and *interpersonal social systems* (Ward & Fortune, 2016b; the RDoC domain of social processes was divided into two separate categories). The X axis contains four different levels of analysis for each of the putative causal processes: *biological, behavioural, phenomenological,* and *contextual* (Ward & Fortune, 2016b; reduced from the RDoC's six units of analysis). The purpose of multi-level data collection is to provide various types of evidence for causal processes, which can then be used to form more comprehensive explanations of risk-related

phenomena (i.e., add explanatory depth). For example, 'emotional congruence with children' may have a number of possible causal processes including a fear/anxiety response to adults, which can be investigated in terms of neural networks and physiological processes, thoughts, beliefs, and emotions, triggering contexts, and resulting behaviours. A virtue of filtering DRF through the DRRF matrix with its core psychological domains and levels of analysis is that their various components and relationships with each other can be more easily discerned. In addition, it avoids the issue of focussing on value-laden and broad domains of DRF because they are teased apart and dispersed across the various functional systems (which are based in human functioning rather than crime).

The DRRF advocates for an understanding of normal or adaptive functioning, in order to better understand dysfunctional processes. The problem is that crime is not necessarily linked with dysfunction; mechanisms could be functioning as intended but directed towards maladaptive or harmful goals (i.e., the normative component). Nevertheless, it is likely that impairments will be present and relevant for a number of individuals, and that explanations pitched at the behavioural and relational (or possibly contextual) levels of analysis can capture the normative components of DRF, while the biological and phenomenological levels deal more directly with facts about the integrity of these systems. Therefore, I suggest that this approach has utility in guiding the theoretical exploration of DRF once they have been broken down into their causal, contextual, and symptom-like (behaviours and mental states) aspects.

6.1.4. Agency impairments.

The fourth approach to the reconceptualisation of DRF rests upon the concept of agency, the capacity for and process of goal-directed behaviour. Recent theoretical papers attempting to link DRF with offending have highlighted the importance of agency and the associated view of offending as goal-directed behaviour (Heffernan & Ward, 2015, 2017; Serin et al., 2016; Thornton, 2016). This idea was developed in the previous chapter. Stressing the importance of focusing on actions as well as underlying mechanisms, Ward, Wilshire, and Jackson (2018) comment "in the context of forensic psychology, our primary goal is to generate etiological explanations of behaviour. Since the targets of our explanations are complex behaviours, the psychological level of description has a privileged status here" (p. 199).

As outlined in the previous chapter, from an agentic perspective DRF are broad categories referring to (i.e., are proxies for) impairments in the capacities underlying agency (e.g., emotion, beliefs, desires, planning, counterfactual thinking, expectations, etc.) and/or social circumstances (e.g., gangs, poverty, unemployment) that lead to behaviour that is

harmful and/or illegal in particular contexts. This means that DRF are contextually bound; what may be a strength or weakness for one individual or context may not be for another. While different theories infer diverse psychological structures and processes, the assertion is that DRF should be conceptualised as problems with the components of agency (intentional, goal-directed behaviour) and/or the contexts in which it is exercised. For example, intimacy deficits could be partially caused by fear responses to adults due to impaired theory of mind capacities (e.g., 'women are cruel'), and poor problem-solving skills could be due to impairments in generating multiple options for action or difficulties with counterfactual thinking.

Conceptualising DRF as problems with the capacities underpinning agency can begin to overcome the issue of their composite nature, as aspects of DRF are dispersed throughout the agency process. Taking the example of emotional congruence with children, an individual values and is motivated to achieve intimacy (i.e., relatedness and pleasure) but believes that adults will harm him, he is in a situation where he feels threatened and lacks the necessary skills to regulate these feelings and he is in an environment with vulnerable children; these impairments and situational factors interact to cause an offence. The focus on human agency means that the first-person, intentional level of explanation is prioritised. Because explanations tend to focus primarily on behaviour and the psychological and situational explanatory levels, they do not *necessarily* incorporate social, cultural, or biological aspects well and thus need to be supplemented with explanations spanning these additional levels of analysis. They are descriptions of action sequences and as such are useful for identifying salient patterns to analyse further and informing the critical explanatory targets of theories of crime.

6.1.5. Exemplars and social practices.

The final way of analysing DRF to be discussed here is using exemplars and normative social practices. As discussed in chapter four, the basic idea is to embed the description of DRF within their relevant temporal and social contexts. The difference between agency models and the exemplar/practices approach introduced in chapter four is that the former is focused on what *persons* do while the latter is on *problems* and their manifestation. In the forensic/correctional context, exemplars could be prototypical characteristics, behavioural patterns, and offence action sequences, such as carefully planned 'grooming' behaviours and the sexual abuse of children. These are reflected within DRF categories and can point to salient features of different types of sexual offending (e.g., preferential, incest, opportunistic; Ward & Siegert, 2002).

Practices are co-ordinated sets of actions centred upon certain goals and their associated norms (Ward & Heffernan, 2017). Norms are evaluative in nature and spell out whether or not an activity is done properly; whether it meets the socially accepted relevant standards. Practices typically depict normative behaviour and DRF (and exemplars) represent violations of these norms. For example, in the case of normative sexual behaviour partners *should* be cognitively competent adults, sex s*hould* occur in private settings, and *ought to* only involve sexual behaviours that are agreed to and are relatively *harmless.* In the case of intimate romantic relationships; they *should* only occur between consenting adults, *ought* to be reciprocal, *should* include personal disclosure of fears and needs, *ought* to incorporate caregiving and sexual components, *should* contain shared activities and responsibilities, and so on.

The practices and exemplars approaches provide a useful way to identify the relevant norms and social/cultural models which govern human behaviour, including norm-violating patterns of behaviour (i.e., illegal practices), which rely upon the functioning of *agency capacities.* It respects the first-person perspective in providing unique insight into intentional practices, but also makes room for a third person (i.e., an observer) perspective in the form of norms and social expectations. Thus, it deals with both the *normative* aspects of DRF, as well as assuming the external conditions for agency are present. However, this approach to reworking DRF lacks depth concerning the structures and processes underpinning agency. The intentional level of explanation cannot tap into the range of sub-personal mechanisms which underlie goal-directed practices. Like the agency approach above, this needs to be supplemented with a more in-depth exploration of causal processes across additional levels of explanation.

6.2. The Risk-Causality Method

Each of the above potential ways of reworking DRF to explain norm-violations and to guide treatment has its own list of strengths and weaknesses. While they propose that the deconstruction and investigation of possible causal elements should play an important role in directing research into the causes of norm-violations, they fail to provide sufficient guidance to researchers and practitioners. The *risk-causality method* (RCM) conceptual framework capitalises on the strengths of the above frameworks by dividing the analysis of DRF into three phases: *deconstruction, analysis,* and *reintegration.* In the RCM, each of the above suggestions for transforming DRF into possible causal elements plays a valuable role, albeit in different phases of the model. The specific theories and models used in this section should be viewed

only as examples to illustrate the utility of the RCM and ought not to be regarded as the correct way to unpack each phase. What is unique is the RCM framework, not the specific models and ideas associated with each of its three phases. Future use of the RCM needs to draw from a greater variety of theories supported by multi-disciplinary research. I will now outline the three phases of the RCM in general terms (see figure 4 below) before applying it to an example DRF.

Phase 1: Deconstruction Identify candidate DRF, exemplar, breakdown into: 1. Symptoms 2. Contexts 3. Possible causal strands (i.e., DRRF causal processes) Output: possible causal processes, multiple for each candidate DRF	**Phase 2: Analysis** Analyse causal processes, acquire evidence across DRRF levels: 1. Biological 2. Phenomenological 3. Behavioural 4. Contextual Output: multi-level local theories for each causal process	**Phase 3: Reintegration** Reintegrate local theories within agency model: Illustrates the role of each causal process in norm-violation, integrates these impairments with normative practices and contexts. Output: a model of behaviour with local theories integrated within

Figure 4. The risk-causality method.[7]

6.2.1. RCM phase one: Deconstruction.

The first phase of the RCM logically follows previous research which regards DRF as *useful markers* of (or as red flags for) crime related factors. Their role in the inquiry process is to direct attention to potential *causes*, relevant *contextual* features, and salient *behavioural* and *mental state* variables (i.e., symptoms). It is important during this first step of deconstructing DRF to view them within the context of goal-directed practices, and to consider whether they could potentially meet Bradford-Hill's (1965) causal criteria. This will ensure that both the factual and normative aspects of DRF are considered. For example, Mann et al.'s (2010) list of empirically supported DRF for sexual offending are a useful starting point. While they do not meet all of Bradford-Hill's (1965) criteria, they are currently the closest thing we have to psychologically meaningful constructs which are linked with this particular type of norm-violation (Mann et al., 2010).

---

[7] Reprinted from "Dynamic risk factors and explanation: The risk-causality method," by R. Heffernan, T. Ward, S. Vandevelde, and L. Van Damme, 2019, *Aggression and Violent Behavior, 44*, p. 53. Copyright 2018 by Elsevier Ltd. Reprinted with permission.

Once a candidate DRF is identified, researchers and practitioners should generate an exemplar (i.e., a typical description) of problems within this category; these will anchor the first phase. Researchers can then describe the behavioural or mental state (i.e., self-report) symptom-like aspects of the DRF, the contexts in which these are observed, and the range of potential causal processes underpinning these. The identification of the different types of constructs contained within each DRF makes it easier to think about the potential causes of vulnerabilities and the way these interact with contextual variables and each other. In order to provide structure to the identification of these causal processes, researchers should consider the six systems listed in the DRRF above; negative affective systems; positive affective systems; cognitive systems; interpersonal social systems; self-regulatory systems; and intrapersonal social systems. Carefully identifying the types of causal processes referred to by DRF will help researchers constrain their subsequent analyses of the relevant mechanisms.

6.2.2. RCM phase two: Analysis.

The second phase of the RCM begins with the list of promising causal candidates from phase one. In the analysis phase researchers should refer each of the possible causes to something like the DRRF matrix to discern their possible *specific causal components* and the evidence for them at different levels of analysis. This process draws from psychological, social, and neuroscientific theories to infer possible mechanisms and processes. While this is potentially the most complicated and time-consuming phase, it is a very important one, and requires thorough investigation in order to provide comprehensive multi-level accounts of the phenomena identified in phase one.

The result of the second phase will be local models of *each* potential causal process (repeated for all of the DRF). These local models should spell out how a system ought to function, and thus be able to identify whether it is operating as it should. In addition, these local models will require reintegration with a more general conception of human agency (in the next phase) in order to understand their interaction and influences on behaviour – no system on its own can explain serious norm violations. The goal at this point is simply to identify the possible causes embedded within DRF and *not to* establish their empirical adequacy; it is a conceptual task not an empirical one.

6.2.3. RCM phase three: Reintegration.

In the third phase of the RCM the goal is to knit the information provided in the earlier two phases together. The various models and relevant contextual features, mental state and behavioural variables are integrated within a model of goal-directed practices, such as the

predictive agency model (PAM; Heffernan & Ward, 2017). This step illustrates the role of causal processes in problematic engagement in normative practices, such as intimacy, sex, coping, and so on. In this sense the final step is reintegrating information concerning the various problems or impairments (the causal processes) underpinning norm violations, with the sorts of practices and environments in which they manifest. The depiction of the RCM so far has been very abstract, so I will now go through each of the phases with an example.

6.3. Illustrating the Risk-Causality Method: Emotional Congruence

In this section I will outline the application of the three phases outlined above to the ongoing example of emotional congruence with children (often listed as a facet of relationship style or intimacy deficits; see Mann et al., 2010; Thornton & Knight, 2015). This will illustrate the use of the RCM in guiding theoretical research into the mechanisms underlying DRF. I make a distinction between the theoretical task of developing general explanations of DRF and the more specific task of individual case formulation. This chapter is primarily focused on the first task but has implications for the second (these will be discussed further in the following chapter). Once the initial theoretical work has been completed in line with the RCM, simplified versions of the models constructed can be utilised in practice with individual cases if and when they are deemed relevant.

6.3.1. Illustration phase one: Deconstruction.

Our example DRF, emotional congruence with children, is one of Mann et al.'s (2010) psychologically meaningful risk factors. Thus, it is reliably linked with sexual reoffending and is considered a *plausible* causal factor. It involves an individual experiencing intimate relationships with children as more emotionally satisfying than relationships with peers. This individual could find children easier to relate to and/or still feel like a child himself and believe that children understand and accept him more than other adults. This often results in feeling like he is 'in love' with his young victims and thinking of the relationship as reciprocal and mutually beneficial (Mann et al., 2010). An exemplar may read: *He is unable to engage in the normal social practices of adult romantic relationship establishment, maintenance, and repair. This is because he values intimacy and sex with children positively and devalues or fears that with adults. He feels safe and valued around children and believes that his victim was a willing participant in the relationship.*

From here it is possible to deconstruct emotional congruence with children into different types of causal processes and contexts/situations that interact to generate the

symptoms (i.e., behaviours and mental states) evident in individuals who feel emotionally drawn to children:

- *Mental state/behavioural variables (i.e., symptoms)*: For example, feelings of fear, loneliness, desire, safety, excitement, hopefulness, despair, avoiding intimacy with adults, making statements reflecting beliefs about children as capable of consenting to and enjoying sex (i.e., willing participant), reporting being in love with victims, using 'grooming' strategies to establish trust.

- *Contexts/situations*: For example, having friendships and intimate contact with children, social isolation, lack of intimacy with adults, regular unsupervised access to children, deviant social networks.

- *Possible causal processes*:

  - *Negative affective systems*: For example, fear and anxiety (i.e., fearful/avoidant attachment); loneliness; guilt/shame (i.e., post-hoc rationalisations).

  - *Positive affective systems*: For example, views children as sources of reward and more likely to signal opportunities for love, sexual pleasure, and care.

  - *Cognitive systems*: For example, displays attentional bias towards signs of affection from children or indicators that it might be possible to become involved with them sexually and emotionally (e.g., cues signifying vulnerability, lack of supervision); beliefs/schema support associations between children and sex (i.e., it is not harmful), developmental deficits (i.e., cognitive impairments).

  - *Intrapersonal social systems*: For example, a tendency to view himself as vulnerable and unsafe, living in a dangerous world; lacks understanding of his motives due to expectancy and interpretational biases.

  - *Self-regulation systems*: For example, lacks the capacity to soothe himself and effectively control negative physiological arousal; seeks interaction with children to do this and constructs elaborate grooming strategies to accomplish this based on problematic beliefs and goals.

  - *Interpersonal social systems*: For example, theory of mind impairments (i.e., inability to take the perspective of and represent others' mental states), internal working models in which affiliation seeking strategies are entirely directed towards daily interactions with children, including sexual contact. These may

be strengthened via involvement in deviant social networks that approve of sex and intimacy with children.

Thus, in this first phase DRF are viewed as *instigators of inquiry* (markers of causal processes), rather than as endpoints of inquiry. The causal processes identified during phase one are the targets for analysis in phase two and are then reintegrated within a model of human agency in phase three.

6.3.2. Illustration phase two: Analysis.

The second phase involves an in-depth analysis of the *potential causal processes* identified, and use of the DRRF (Ward & Fortune, 2016b) to investigate each of these systems across multiple levels of analysis: biological, behavioural, phenomenological, and contextual. Due to limited space, I will focus on the example of a negative affective system identified in phase one above; the tendency to view adults as threats and a source of fear and anxiety, resulting in avoidance of intimacy with adults and a preference for intimacy with children. At the *biological level*, this system could involve the amygdala, central nervous system, and associated physical responses to fear and anxiety (e.g., heart racing, dry mouth, and perspiration). At the *behavioural level* this system is primarily concerned with fight or flight responses, for example avoidance of interactions with adults or hostility and mistrust towards them. These behaviours could manifest in a lifestyle (i.e., social and physical contexts) characterised by social isolation or intimacy deficits and seeking out children as a safer alternative. The contexts in which these systems are activated may include social events where unknown adults are present, rejection from adults, and where norms require social interaction. At the *phenomenological level* this system involves beliefs about adults (e.g., rejecting, dangerous, manipulating), memory and attentional biases (i.e., towards events that confirm these beliefs), and associated emotions such as fear, anxiety, and loneliness.

In order to gain a deeper understanding of how things have gone wrong, researchers need to explain how a threat detection system *ought to work*. Its functioning likely exists upon a continuum, with a healthy range existing within the middle, and sub-optimum functioning when the system is over- or under-functioning. It is also context-dependent and normatively defined, for example it is normal and adaptive to be more sensitive to interpersonal threat in prison, whereas intimacy practices require a certain level of trust and vulnerability. For this reason, the contextual level must be explored and integrated with lower levels of analysis. The analysis phase involves integrating evidence from a number of disciplines, for example

psychopathology (e.g., social phobia), evolutionary psychology (e.g., the selective advantage of fear responses), developmental psychology (e.g., attachment), and neuropsychology and biology will be particularly useful for lower level explanations. This work remains to be done and is outside the scope of this thesis. It is anticipated that the output from phase two will be a number of local models centred upon each identified causal process and spanning multiple levels of analysis.

It is also worth noting at this point that the functioning of the negative affective system (and others) is relevant to a wider range of symptoms and behaviours which reflect the violation of accepted social norms, not just emotional congruence with children and sexual offending. For example, dysfunctional threat detection may play a role in social anxiety, general relationship problems, violent offending, and substance abuse. The same underlying mechanisms can produce different behaviours (both legal and illegal) depending upon the context and the functioning of other systems. It may lead to sexual offending against children when paired with a sexual preference for children, beliefs about the ability of children to consent to sexual intimacy, or self-regulation problems (i.e., intoxication lowering inhibitions) in certain contexts.

6.3.3. Illustration phase three: Reintegration.

The third phase involves the reintegration of causal processes and their associated systems (i.e., local theories) within an *agency framework* containing emotional capacities; psychological representations; cognitive processes; formulating goals and plans; implementing plans; evaluating outcomes; and modifying plans and goals within day to day activities (i.e., goal-directed practices). The goal of this phase is to provide an understanding of how the putative causal processes contained within each DRF interact with each other and contextual features to produce offending behaviour. For illustrative purposes, we will embed our example within the previously developed PAM (Heffernan & Ward, 2017).

In this (speculative) example, underlying fear/anxiety towards adults are general models of the self as vulnerable, other adults as dangerous, and children as safe and accepting. In terms of affect, the individual is lonely due to avoidance of peers, and is overly sensitive to interpersonal threat. In social situations, such as family gatherings, he tends to pay attention to social cues which support his internal working models, and construct situation-specific models which are in line with these. For example, if certain adults do not talk to him it is perceived that they dislike and want to harm him. He has developed the ability to quickly create friendships with children and get parents to trust him (i.e., he can anticipate their reactions to

his interest in children). He has excellent planning abilities and is always thinking several moves ahead; he rarely acts impulsively.

I have taken just one example of a DRF and used the RCM to tease out possible causes, contexts, and behavioural and mental state variables (i.e., symptoms) in order to explain how they might manifest in harmful behaviour. It is clear from this example that emotional congruence with children is linked to other DRF, such as problematic beliefs and deviant sexual preferences. An advantage of breaking down DRF and embedding them in behaviour in this way is that it becomes easier to formulate cases. Once further theoretical work has been completed, practitioners will have access to a *causal model* not simply a *statistical/predictive one*. This offers greater potential for accounting for diverse behavioural features with an integrated set of mechanisms, is more individualised, and has greater explanatory depth – it provides more information than a formulation stating that "this individual emotionally identifies with children".

6.4. Conclusions: Risk-Causality Method

I have argued throughout this thesis that DRF should not be accepted at face value as possible causes of offending because of the problems of incoherence, lack of specificity, the grain problem, and their normative status. In this chapter I introduced the RCM with its phases of deconstruction, analysis, and reintegration as one promising way forward in the investigation of the potential causal elements contained within the broad DRF categories. I propose that the RCM can be usefully applied to DRF to 'boot strap' theory development and eventually provide a valuable source of information for formulating cases. At this stage it is intended primarily as a research model which aims to provide a link between theory and practice.

A number of hypotheses follow from the RCM, I will outline three of the more interesting ones. First, I hypothesise that each DRF, as formulated in the risk prediction and management literature is constituted by a combination of *different types* of causal processes. For example, positive and negative valance systems, different types of cognitive processes, interpersonal processes, and self-regulation processes converge to constitute a DRF such as self-regulation difficulties. This means that formulating DRF as primarily social, affective or as cognitive in nature is misleading. Second, DRF can be described at *different levels of analysis* and the processes evident at each level mutually constrain each other. Furthermore, individuals will vary in terms of what levels are primarily responsible for generating the problems reflected in DRF, and ultimately in problematic behaviours such as crime. Finally,

treatment based on RCM-oriented case formulation will be more effective than the current practice of simply targeting DRF as described in risk prediction instruments.

This chapter has concentrated on the role of DRF in explaining offending and its associated problems. However, the RCM can also be utilised with other types of potential causal factors, such as those contained in etiological theories or implicit within classification systems. For example, attachment models of intimacy deficits in those who have committed sexual offences propose that problematic social behaviour is in part caused by impaired internal working models of relationships (Smallbone & Dadds, 2000). Maladaptive core beliefs of the self and other people are aligned with interpersonal strategies designed to protect individuals from rejection and overwhelming negative emotional states. Alternatively, according to the *self-regulation model* of the offence process (Ward & Hudson, 1998b), individuals who commit sexual offences vary in terms of their core values, goals, and strategies for managing problematic desires and situations. These differences manifest in distinct patterns of offending behaviour (Yates & Kingston, 2006). In both examples, the RCM could be applied to possible causal mechanisms such as impaired internal working models or self-regulation styles to develop richer theoretical depictions of them. In addition, researchers should trace their links to various mental states, behaviours, and contextual features, and spell out any possible treatment implications. Thus, the RCM has the potential to unify theory development tasks in a variety of research domains and has a much wider reach than simply breaking down DRF. The RCM may assist researchers and practitioners to bridge the gap between theory and intervention in the criminal justice system. In order to intervene to effectively reduce crime it is necessary to move beyond risk assessment and management. Ultimately this requires us to appreciate how DRF might tap into the mechanisms which relate to offending and understand how best to ameliorate their influence.

Chapter Seven: Practice Implications of a Focus on Agency Mechanisms

The usability and utility of theory in practice should be of key concern to researchers. While it may be satisfying and valuable to develop a coherent and comprehensive explanation for certain phenomena, ethical concerns dictate that researchers should also consider the practical implications of their work. At this point I would like to point out that despite some promising findings, the current state of correctional intervention is lacking and suggest that we need to do better. As summarised by Hough (2010):

> The most consistent finding to emerge from those examining programmes for reducing offending is that insufficient research has been conducted of high enough quality to say much with any confidence. The second most consistent finding… is that knowledge about what works is inconsistent and incomplete. The third finding that consistently emerges is that some programmes sometimes work. Those that have emerged as successful most frequently are cognitive behavioural programmes. This probably reflects the fact that they have been developed on a firm theoretical foundation – but also that in their nature, they are amenable to evaluation. (p. 13)

I suggest that at least some of this inconsistency is due to confusion surrounding its foundational constructs, DRF. In line with the need principle (Andrews & Bonta, 2010), DRF are the targets of effective interventions. They guide tasks such as intervention assessments and case formulation, treatment planning and goal setting, developing programme manuals, and the measurement of progress. Throughout these tasks, theories are used by practitioners to understand their clients' presenting problems and work with them to facilitate change and observe progress against areas of need (Ward, Polaschek, & Beech, 2006). Without a coherent understanding of what DRF are, how they relate to offending, and how they might change, practitioners are limited in their ability to promote change and reduce harm.

This chapter will explore what could change at present to both acknowledge the limitations of our knowledge concerning the causes of offending and capitalise on what is known about human behaviour. While significant theoretical work remains, there are implications for viewing DRF (and offending) as aspects of goal-directed practices, normative contexts, and predictive agency capacities. I will first briefly outline the current approaches to criminogenic treatment, and comment on the role of values within rehabilitation models. I will then discuss potential issues with the standard reliance on evidence primarily from treatment

outcome studies to inform our understanding of whether and how rehabilitation works. I suggest that a combination of research techniques might better inform interventions; we must widen the scope of what we consider evidence for what works. I will then outline the *good lives model* (GLM; Ward & Maruna, 2007) and discuss how it aligns with a view of persons as agents who engage in goal-directed practices. Finally, I will make some suggestions for how practice might look moving forward, building upon the RNR, GLM, and the reconceptualisation of DRF advocated for throughout this thesis.

7.1. Rehabilitation Theories and Values

It has been proposed that rehabilitation frameworks contain 1) overarching values, principles, and aims, 2) assumptions about the causes offending and behavioural change, and 3) practice guidelines, principles, and tools for use in interventions (Ward, Melser, & Yates, 2007). For example, the RNR model aims to reduce victimisation (reflecting social and ethical values) by taking a 'rational empirical' approach to research (reflecting epistemic values); it contains general principles (i.e., RNR), is based on a *general personality and cognitive social learning* view of persons and the assumption that DRF cause offending, and contains additional principles and risk assessment tools to guide practice. The RNR model proposes that effective intervention occurs via change in criminogenic factors through collaborative, compassionate, and dignified human service (Polaschek, 2012). However, in practice responsivity to the individual and their unique needs is often overlooked (Polaschek, 2012). Needs which are not empirically linked with reoffending are largely seen as a responsivity issue – they may be targeted if they are barriers to treatment. The GLM augments the RNR by providing an approach to treatment, rather than a set of guidelines for correctional practice generally. It offers a strength-based view of persons and views offending behaviour as motivated by normative human needs. This is compatible with the RNR, as it also views risk reduction as a core goal; it simply disagrees that an exclusive focus on DRF is the best way to achieve this.

I suggest that an important but often unappreciated aspect of rehabilitation theories is their reliance upon values and the social and normative aspects of the offending change process. It is clear that the decision to fund certain types of intervention or research into crime is in part based on the social and ethical priorities of the community, for example reducing victimisation. However, these funding decisions and priorities are also based upon epistemic values which dictate the types of knowledge and evidence which are considered sufficient to guide practice. For example, the widespread reliance on meta-analyses and treatment outcome studies which have provided information about the correlates of recidivism and the treatment modalities

which have demonstrated the most success in reducing aggregate rates of recidivism (i.e., the PCC and RNR). It is concluded from this research that risky or antisocial aspects of the person (i.e., correlates of recidivism or criminogenic needs) must be 'fixed' via cognitive-behavioural interventions. Interventions designed to improve the overall functioning or wellbeing of those who have committed crime are not viewed as a priority, or at least they are not the core business of Corrections (Bonta & Andrews, 2017; Polaschek, 2012). While this may appeal to some as it aligns with punitive ideals, it does not necessarily reflect the most effective approach to rehabilitation and thus may not support the aims of the justice system.

7.2. The Risk-Need-Responsivity Model

As the RNR model was outlined earlier in this thesis, I will therefore only briefly repeat the core principles here. The risk principle states that individuals at higher risk of reoffending will benefit most from more intensive intervention and that lower risk clients should receive less intervention. The need principle proposes that changeable features reliably associated with reductions in recidivism (i.e., DRF or criminogenic needs) should be targeted in treatment programmes relative to those that have no demonstrated empirical relationship to crime. The responsivity principle states that correctional programmes should use empirically supported treatment models (e.g., cognitive-behavioural therapy) and be responsive to individual characteristics such as learning style and level of motivation. The first two principles (risk and need) are used to select treatment intensity and targets, and the whole set of principles are employed to guide the way practice is implemented.

7.2.1. The RNR and Values.

Reflecting their legal and ethical origins, in my view the RNR principles are partly normative in nature rather than simply being empirical laws or generalisations. The assumption is that failure to follow these principles is likely to result in more reoffending, which would be poor practice, irresponsible, a bad thing, unethical, irrational, and so on. In addition, the risk principle is reliant on value judgments specifying the meaning of the various risk bands of low, medium, and high. The key normative question is how should communities weight the likelihood that individuals might offend in the future against their own concerns (i.e., whose interests should they privilege and why?). Furthermore, rehabilitation programmes directly aim to change the way participants engage in day to day activities. Actions are norm-governed (i.e., they are aimed at desired outcomes and evaluated against standards of performance), and norms are explicitly based on values (i.e., the worthiness of the outcome and the success of the action

with respect to the goal). Therefore, all RNR programmes are value-laden; they provide guidelines for intervention which are based on normative concerns and practices.

While the RNR principles are based on empirical research, as normative principles they are not factual statements. Therefore, they are not directly empirically confirmable or capable of falsification because they are prescriptive guidelines for practice. It would be misleading to say because they have not been falsified – or are based in evidence – that the RNR has been strongly supported as a theory. Such principles are not logically capable of falsification; they are simply useful practice guidelines, or they are not. This clarification points to the complex nature of rehabilitation theories in the forensic and correctional area (i.e., comprised of values, empirical, and practice components). Because of considerations such as an appeal to possible harms and the obligation to reduce reoffending, social and ethical values are embedded deeply within the RNR principles. Arguably this reliance on norms and values is unavoidable when dealing with normative constructs such as crime, harm, risk, and antisocial traits. However, these elements should be acknowledged and kept separate from claims about evidence and causal mechanisms. The authors of the RNR have not clearly separated out the causal from the normative aspects of the model, and this is evident in their rational-empirical approach to understanding and treating the causes of crime.

I have concerns with this model which stem from its prioritisation of prediction over explanation and its tendency to rely heavily on inconsistent evidence which does not adequately support its assumptions about the causes of offending (see chapter two). In terms of its core principles, these issues mainly impact upon the *need* and *responsivity* principles. The need principle is based on statistical evidence concerning the best predictors of recidivism and assumptions about the (possible) causal nature of DRF. As argued throughout this thesis, DRF in their current state cannot be considered causes of crime. The general responsivity principle requires that practitioners and programme designers are responsive to what works, and I argue that we need to look beyond treatment outcome and studies and variation in recidivism rates in order to understand whether and why certain types of treatment are successful. Finally, the specific responsivity principle requires that practitioners are responsive to the needs of individuals. I argue that this requires both research into individuals' experiences of change and rehabilitation, as well as a comprehensive understanding of human functioning and agency. I will now explore each of these issues in more detail.

7.2.2. The focus on recidivism and statistical associations.

Programmes adhering to the RNR principles have demonstrated *relative* success (compared to those which do not) in reducing recidivism rates (Andrews & Bonta, 2010; Bonta & Andrews; 2017; Dowden & Andrews, 2000; Gannon, Olver, Mallion, & James, 2019; Hanson, Bourgon, Helmus, & Hodgson, 2009) and so long as this continues, they should be used to guide practice. For example, Hanson et al. (2009) reported 10.9 per cent sexual and 31.8 per cent general recidivism following treatment for sexual offending (versus 19.2 and 48.3 per cent in the untreated group, respectively) and found that programmes adhering to the RNR principles had greater reductions than those which did not. The body of work which supports these principles is currently the most robust evidence available to guide practice, and I am not suggesting that we abandon all we currently know about what works. However, as pointed out by Hough (2010):

> … there has been over-investment (both financially and intellectually) in a technocratic model of reducing reoffending that attaches too much importance in accredited programmes and packages, and under-investment in models that see the process of 'people changing' as a complex social skill. The technocratic model seriously underestimates this complexity, and its advocates wrongly assume that experimental research can readily identify the causal processes at work in helping people to stop offending. (p. 12)

Statistical associations between exposure to intervention and rates of recidivism offer a general (but imperfect) answer to the question of whether treatment 'works' but do not tell us whether and how or why people have made enduring changes to their lives. Some limitations with treatment outcome studies were discussed in chapter two, these included various methodological weaknesses, the subjectivity of risk ratings and reliance on self-report data, the composite and eclectic nature of treatment programmes, and the use of recidivism as an outcome measure.

I argue that these studies offer a rather narrow view of the available knowledge concerning rehabilitation, desistance, and change, and suggest a broader interpretation of what constitutes evidence and success. There are three reasons for this stance: 1) recidivism is a difficult to measure and decontextualised outcome variable (Jung & Gulayets, 2011), 2) the evidence concerning criminal behaviour ought to be drawn from multiple disciplines and inform explanations of individual functioning rather than relying upon lists of correlates, and

3) the success of current rehabilitation programmes has not yet been satisfactorily proven and we need to consider new ways to increase their effectiveness. Knowledge is cumulative, rather than one definitive study proving that treatment is effective or ineffective, and our understanding of its effects will grow through the accumulation of many smaller studies using various methods (Collaborative Outcome Data Committee [CODC], 2007). In relation to this issue, Farrall (2005) makes the point that:

> … without a willingness at least to consider in-depth the experiences of individuals who have successfully negotiated the transitions from "offender", it is unlikely that efforts to encourage desistence (e.g., the What Works movement) will produce the sorts of results so desperately needed. (p. 383)

This mainly relates to the general responsivity principle (i.e., how we should go about treatment in an evidence-based way), but also has implications for the other principles. In order to be truly responsive to the evidence concerning what works, we must broaden the scope of the research relied upon to inform intervention design (i.e., general responsivity) and delivery (i.e., specific responsivity). This means paying attention to qualitative research and listening to participants' perceptions about what they need in order to change. It also means developing an understanding of human behaviour which is not based solely on the correlates of crime. By focusing on the way systems ought to function in order to meet certain needs or perform certain tasks, researchers have a better chance of determining what has gone wrong when these processes lead to offending.

7.2.3. The reliance on correlates as causes: The need principle.

The analysis of DRF provided throughout this thesis has seriously called into question the legitimacy of relying upon these concepts as explanations for offending and targets of interventions. This is the basis of the need principle, so I suggest that this principle is either fundamentally flawed or at best under-developed. Further, given the reliance of the other RNR principles on the concept of DRF, the theoretical basis of the whole model is arguably undermined. It is likely that the domains of functioning and arenas of life reflected within DRF categories are in 'need' of prosocial change (i.e., in order to reduce harm), or that when psychological change occurs, we will also see changes in these areas (i.e., they are indicators of a criminal or prosocial life orientation). However, without an understanding of *what* needs to change and *how* this might happen, we are unable to adequately plan and deliver treatment or monitor its success. Useful case formulation requires that practitioners understand the

probable *causes* of the problems reflected within criminal lifestyles and patterns of behaviour. In addition, it is worth considering that the factors which reduce crime may be different to those which have been shown to precede offending (Kroner et al., 2017; Polaschek, 2016; Serin & Lloyd, 2009). In other words, even if we understand the causes of offending, we may then need to also look to the causes of desistance in order to understand and influence the change process (this will be further discussed below in relation to GLM programmes). One potential way to address this issue is to adopt a more holistic view of persons as goal-directed agents motivated by universal needs.

7.3. The Good Lives Model

As discussed above, the GLM broadens the focus of rehabilitation from simply the reduction of risk to also providing participants with the resources required to live a good life. I will now outline the GLM in more detail, as it provides a treatment approach which I will build upon with my suggestions for practice. The reason for this focus on the GLM is that it augments the RNR model by providing a motivating and personally meaningful approach to therapy that fits well with an agency perspective of human behaviour. The approach to DRF developed throughout this thesis is consistent with the GLM, and the PAM adds depth to the model by specifying the components of agency. Following my description of the GLM and the contributions of this approach, I will describe this addition in more depth.

The GLM is a strength-based approached designed to enhance the RNR model, rather than act as a replacement or alternative to the model. As discussed above, the GLM's view of persons is as goal-directed beings who are motivated to engage with their environments in order to pursue a range of valued outcomes, termed 'primary human goods' (PHG). The PHG contained within the GLM have varied slightly over time, a recent list includes excellence in work, excellence in play, creativity, knowledge, relatedness, community, pleasure, life, inner peace, spirituality, and excellence in agency (Barnao, Ward & Robertson, 2015; Purvis, 2010). The means (i.e., behaviours, strategies, outcomes) which persons use to meet these needs are termed 'secondary human goods'. For example, one person may meet their need for 'inner peace' (i.e., freedom from emotional distress) through regular exercise or therapy, while another may use substances or violence to deal with emotions. The purpose of treatment according to the GLM is twofold; to prevent potential harm to victims (i.e., reduce reoffending) and provide the individual with the internal and external resources required to live a personally fulfilling good life. Proponents of the GLM suggest that these two aims are not mutually exclusive, as criminal behaviour is seen as a consequence of problems within an individual's (largely implicit) 'good life plan' (GLP). This plan contains both primary and secondary goods,

PHG are prioritised to different degrees across individuals and situations, and the secondary goods used depend on both opportunities available and individual capacities. Thus, according to the GLM, the causes or contributors to offending and related problems (i.e., DRF) are deficits or barriers in the internal and external resources needed to achieve PHG in ways which do not harm others. These problems should be the focus of treatment; interventions aim to build strengths in these resources in order to meet the dual aims of risk reduction and good life promotion.

According to Ward and colleagues (Ward, 2002a, 2002b; Ward & Fisher, 2005) there are four types of problem which can arise and often interact or co-exist within GLP: capacity, means, conflict, and scope. In brief, when *capacities* or resources (internal and external) are lacking, an individual may turn to antisocial behaviours (i.e., harmful *means*), such as offending, in order to attain PHG. Sometimes problematic actions and their outcomes can *conflict* with an individual's attainment of other goods. For example, poor coping abilities or lack of problem-solving skills (capacity) may lead to use of violence (means) aimed at attaining excellence in agency or control, which may result in imprisonment and obstruct an individual's ability to meet needs such as relatedness or pleasure (conflict). The final problem, lack of *scope,* occurs when certain goods are prioritised at the expense of others, and not all are represented within a person's GLP. In addition, GLM researchers have expanded upon these four problems to empirically identify and describe two routes to offending, *direct* and *indirect* (Purvis, Ward, & Willis, 2011). The direct route occurs when an individual's capacities and resources are lacking, and offending is used as a means to attain PHG. The indirect route involves conflict or problems with scope, and offending occurs as a result of other unhelpful behaviours rather than being directly used to meet needs.

7.3.1. The GLM in practice.

GLM interventions are based upon a participant's personally meaningful GLP and developing a future which does not involve harming others. The GLP contains all PHG categories to varying degrees (chosen by the individual) and the secondary goods (both goals and strategies) required to attain these in healthy and prosocial ways. It is important that the dual focus on reduced risk and good lives is evident throughout assessment and treatment. For example, these two priorities can be linked by identifying which PHG may have been related to offending (directly or indirectly) and exploring these relationships – for example, are there problems with capacities, means, conflict, or scope? Many of the problems identified are likely to correspond to established DRF categories, for example, capacities such as problem solving,

emotional-coping, and/or perspective-taking may be lacking. Treatment goals will include strengthening these capacities (internal resources) so the individual can meet their needs without relying upon harmful means. In addition, external resources and conditions (i.e., opportunities and constraints) can support or obstruct goods attainment, and so these should also be included in a comprehensive GLP. In summary, initial assessments should identify an individual's most important PHG and those linked with offending, use these to construct a personally meaningful GLP, and then treatment should aim to develop or strengthen the (internal and external) resources required within this GLP. There is a dual focus on promoting PHG and overcoming barriers (i.e., criminogenic needs, problems within past GLP), this is achieved through collaboration and a focus on individual goals, rather than simply avoiding risk.

The GLM is a strengths-based approach to correctional rehabilitation because it is responsive to people's particular interests, abilities, and aspirations (Ward & Maruna, 2007). It also asks practitioners to explicitly construct intervention plans that help people acquire the capabilities to achieve personally meaningful goals. Rehabilitation plans should aim to equip persons with the knowledge, skills, opportunities, and resources necessary to satisfy their needs and values in ways that do not harm others. Related to its strong focus on persons' core commitments and lifestyles that reflect these, there is a corresponding stress on agency. Because of the assumption that individuals who commit crimes, like the rest of us, actively seek to satisfy their needs through whatever means are available to them, any rehabilitation plan should be pitched at the level of agency, goals, planning, and environments. The GLM is an ecological model which keeps in mind the relationship between the environments in which persons live and the capabilities and resources they need to live meaningful and crime free lives (Ward & Stewart, 2003). Criminogenic needs or DRF are conceptualised within the GLM as internal or external obstacles (i.e., flaws within a good life plan) that make it difficult for individuals to secure primary goods in personally meaningful and socially acceptable ways.

Proponents of the GLM have developed guidelines and practice tools to aid in the integration of the model into programmes (e.g., Purvis et al., 2011; Willis, Yates, Gannon & Ward, 2013; Yates, Prescott, & Ward, 2010). It should be used throughout the tasks of assessment and case formulation, case management and intervention planning, development of programme manuals, and developing the therapeutic relationship. The GLM is used internationally, including in New Zealand (Whitehead, Ward & Collie, 2007), Australia (Willis & Ward, 2013), the United Kingdom (Langlands, Ward, & Gilchrist, 2009), Belgium (Van Damme, Hoeve, Vermeiren, Vanderplasschen, & Colins, 2016), Singapore (Chu, Koh, Zeng,

& Teoh, 2015), the United States (Willis, Ward, & Levenson, 2014), and Canada (Marshall, Marshall, Serran, & O'Brien, 2011). It has been suggested as appropriate for use with a range of participants, including youth (Fortune, 2017; Wainright & Nee, 2014; Wylie & Griffin, 2013), female (Van Damme et al., 2016), elderly (Di Lorito, Vollm, & Dening, 2018), intellectually disabled (Aust, 2010), mentally disordered (Barnao, Ward, & Casey, 2015; 2016; Gannon, King, Miles, Lockerbie, & Willis, 2011), and non-Western (Chu, Koh, Zeng, & Teoh, 2015; Leaming & Willis, 2016) offending populations. In addition, beyond its preliminary use in interventions for sexual offending it has been considered useful for the treatment of substance abuse (Thakker & Ward, 2010), violence (Whitehead et al., 2007), domestic violence (Langlands et al., 2009), and residential burglary (Taylor, 2017).

In terms of its role in these programmes, it is often viewed as an 'add-on' to traditional risk focused interventions, rather than being used as intended to guide the entirety of treatment. For example, Willis et al. (2014) evaluated the operationalisation of the GLM in North American programmes and found that it was most often used within programme delivery (i.e., developing positive therapist characteristics) and as an additional component to treatment (e.g., self-management plans at programme completion). These authors proposed that "enhancing program consistency with the GLM requires using it as a comprehensive theoretical framework to guide interventions throughout the entirety of a program" (Willis et al., 2014, p. 77). Thus, while the GLM is utilised internationally, there are issues with its implementation and inconsistencies with its intended role throughout the rehabilitation process.

7.3.2. GLM and values.

The GLM is explicitly oriented towards the goals of increasing individuals' level of well-being and reducing their chances of harming other people in the future, reflecting prudential and ethical values respectively. The reference to primary goods is strongly naturalistic and grounds agency in the basic orientation of life forms towards sources of reward and away from threats. Furthermore, the pivotal role of a GLP in the rehabilitation process places core value commitments at the centre of any change process; clearly a normative commitment. In its comprehensive formulation, the GLM distinguishes normative assumptions from empirical claims and therefore does not share the same weakness the RNR exhibits in this respect. However, its emphasis on multiple types of prudential goods creates problems related to ranking them in practical decision-making, a task that it does not deal with explicitly. In addition, theorists working on the GLM have not yet discussed how to address conflicts between social and ethical values, and peoples' GLP. However, ethical codes may be somewhat

helpful in resolving these types of conflicts alongside the GLM. For example, in the GLM all individuals affected by crime have equal moral status and therefore their interests should be explicitly considered when creating intervention policies. In accordance with this ethical assumption, the two major objectives of forensic and correctional rehabilitation interventions are to reduce recidivism rates and to increase the well-being of individuals who have committed crimes.

7.3.3. Contributions of the GLM.

As a rehabilitation framework, the GLM provides a strength-based perspective on the nature of persons and guidelines for practice, rather than being a specific programme itself. This means that evidence supporting its use comes from evaluations of programmes consistent with GLM values and assumptions, and their impact on a range of outcomes. Empirical research supporting the GLM is limited in comparison to the abundance of theoretical papers and those outlining its *potential* utility with various populations. As discussed above, the GLM is used in different ways across programmes internationally, and because it is designed to augment the RNR, its use should result in programmes which are *at least* as successful in reducing recidivism. For this reason, the research investigating the success of the GLM tends to focus on what it adds to the standard approach of reducing risk in RNR-consistent programmes. Concepts such as motivation, engagement, and cognitive transformation are crucial in understanding how and why treatment works, and these are best accessed via first-person accounts of programme experiences and progress rather than empirical studies for which recidivism is the outcome variable. Thus, studies investigating the use of the GLM are most often qualitative and take the form of case studies or programme comparisons using relatively small samples sizes. I suggest that this form of knowledge should not be overlooked, as it is no less useful than statistical information derived from large groups, it simply gives us an answer to different questions. For example, we cannot hope to understand the process of change, the meaning attached to it, and what has worked for individuals in building motivation and supporting change by just looking at reconviction rates. Perhaps this is one reason for the under-development of the responsivity principle – unlike the risk and need principles, it does not lend itself to investigation via quantitative methods.

I will now briefly discuss four themes which have emerged from studies incorporating GLM concepts and values: the use of approach goals, prudential values and PHG, specific responsivity (i.e., an individualised and collaborative approach), and programme retention and engagement. This discussion will draw from some of the evidence supporting the GLM as it

relates to each of these themes, and also make some comments about how this relates to the conceptualisation of DRF developed throughout this thesis.

Approach goals. The use of approach goals to motivate and guide treatment is central to the strength-based approach of the GLM. Approach goals are directed towards a desired outcome and developing the internal resources needed to achieve this in healthy and prosocial ways. For example, an approach goal might be to build the capacities and skills needed to establish and maintain a healthy relationship with an adult. In contrast, an avoidant goal might be to stay away from places where children are present. Both are aimed at addressing problems with intimacy and preventing sexual offending, and indeed both might be helpful for certain individuals or times when risk is high. However, the approach goal focuses on positive outcomes rather than just removing sources of risk – what has been described as a "pin cushion" approach (Ward, Mann, & Gannon, 2007). Approach goals have been found to be more motivating than avoidance goals, and their effects have been observed to last longer (Marshall & Serran, 2004). In addition, the use of approach goals in a GLM adapted programme resulted in a higher completion rate, greater motivation (perceived by therapists), participants gaining more social support, and significantly greater improvement on psychometric measures of problem-solving and coping skills compared to an avoidant approach (Simons, McCullar, & Tyler, 2008). A similar study found better engagement (i.e., disclosure and task completion) and therapists perceived more genuine motivation for prosocial change in the approach goal condition (Mann, Webster, Schofield, & Marshall, 2004). Therapists in this study expressed concern that the approach goal programme was more complicated to implement and that participants may lack an understanding of their risk factors (Mann et al., 2004). However, they measured participants' awareness and understanding of risk factors and risk-management strategies, and found no significant difference between the approach and avoidant groups – *both* had significantly improved (Mann et al., 2004). Still, practitioners must ensure that programmes address needs linked to offending and include strategies to manage risky situations alongside promoting the good life through approach goals. Achieving this balance is only possible when practitioners are well trained, have a sound understanding of human functioning and the probable causes of different behaviours, and are encouraged to work in a flexible and responsive way.

Prudential values and primary human goods. PHG are also central to GLM interventions, representing outcomes towards which approach goals may be directed. The GLM is concerned with ethical and social values (i.e., reducing harm), and prudential values (i.e., wellbeing or personal priorities) – these are not mutually exclusive. The concept of PHG is

able to capture the role of prudential values in offending, and thus provide a way to understand and link these dual aims. A number of studies have investigated the role of PHG in offending and the change process. For example, Barnett and Wood (2008) explored participants' experiences at the time of their offending and found evidence of the four types of problems in GLP described above. For example, problems with the prioritisation of goods and operationalising and balancing a GLP (i.e., scope and conflict), and poor problem-solving skills, difficulty in achieving PHG, and use of offending to meet needs (i.e., capacity and means). Other studies have explored differences in PHG across samples and in relation to different offending behaviours. For example, Chu et al. (2015) found that pleasure (91.1 per cent), relatedness (35.7 per cent), and inner peace (17.3 per cent) were most highly prioritised for youth who had sexually offended in Singapore. In another example, Loney and Harkins (2018) found support for the importance of PHG and links with self-reported offending (via problems with capacities and means) in a student population; life, knowledge, and happiness were prioritised most, and agency, inner peace, and happiness were most often sought via problematic means. In terms of behaviours, Loney and Harkins (2018) found links between inner peace, agency, and *violence*, and happiness, inner peace, and *drug offences*. Taylor (2017) found evidence for the relevance of PHG in residential burglaries, for example, participants talked about the 'buzz' (i.e., pleasure) from offending, whereas others spoke about being good at it (i.e., excellence in work/play).

PHG may also vary in terms of their importance across time, for example some may be more relevant to reintegration and others in treatment. Harris, Pednault, and Willis (2017) found that released individuals who were incarcerated for sexual offending valued many PHG, but that their ability to achieve them was significantly restricted by their correctional status. Barriers to meeting needs included a lack of supportive relationships due to offending and the consequences of disclosing past behaviour, and problems securing employment and accommodation. In addition, Willis and Ward (2011) found that positive experiences of re-entry were associated with later attainment of PHG, suggesting that these experiences can facilitate or restrict PHG attainment. Marshall et al., (2011) evaluated a sexual offending programme containing six areas of PHG and found recidivism rates below what was expected (based on meta-analyses) and predicted before treatment. Marshall et al. (2011) concluded that "when programs target problems that are obstacles to treatment, and then focus on changing known criminogenic features by taking a positive, respectful, and process-oriented approach, the re-offense rates… are likely to be significantly reduced" (p. 92). A more recent evaluation of this programme provides further support for a strength-based approach to treatment with

individuals who have committed sexual offences (Olver, Marshall, Marshall, & Nicholaichuk, 2018).

The above findings support the importance of PHG across samples, but also suggest that their prioritisation varies across groups and in relevance for different behaviours (i.e., offending). They also support the assumption of the GLM that the prioritisation of goods and problems in their attainment are linked (directly or indirectly) with offending and offer preliminary support for the idea that PHG attainment may support treatment change and the desistance process. In terms of linking this with agency perspectives, PHG (paired with emotional states) are sources of motivation to act and secondary goods are an individual's strategies associated with these motivators (i.e., contained within general models). These PHG and secondary goods are linked within general models through past experiences of reward and associated emotions (e.g., certain behaviours linked with pleasure or relief). Attainment of PHG is therefore dependent upon the general models and internal resources required for a certain type of practice (i.e., directed at PHG), what I have termed predictive agency capacities.

Specific responsivity. As above, PHG are empirically supported and theoretically grounded universal values (Laws & Ward, 2011). However, they are viewed as being of varied importance for different individuals and are attainable by a range of goals and strategies. The GLM does not make assumptions concerning what a good life should be like; the list of PHG is a guide to expand the scope of an individual's GLP, rather than direct them towards outcomes they do not care about. For example, the PHG 'knowledge' does not require a formal education, but rather that individuals identify the kinds of knowledge they value, or which would support their goals (e.g., vocational, self-knowledge, etc.). This flexible approach is responsive to personal needs and differences in persons' conceptions of a good life. Thus, the GLM is able to overcome the restrictions of a 'one size fits all' approach to rehabilitation, based on lists of problems found at the group level. It is a collaborative approach (Yates & Ward, 2008), which means that it prioritises participant agency and autonomy in treatment planning and the change process. The importance of interpersonal factors in treatment is widely acknowledged, and I propose that the GLM encourages interactions that are respectful, warm, non-judgmental, and engaging. Indeed, research has found that in sexual offending treatment for youth the GLM "appears to impact positively on the therapeutic alliance, promote self-efficacy and optimism and increase the client's capacity to succeed and address issues of risk" (Wylie & Griffin, 2013, p. 354).

Retaining and engaging participants. Treatment drop-out or attrition represents a threat to the effectiveness of programmes. Those who begin but do not complete treatment tend

to have higher rates of reoffending than those who do complete treatment, and possibly also those who do not receive treatment at all (see McMurran & Theodosi, 2007). There are many reasons why individuals fail to complete treatment, including voluntary exit, rule breaking, and administrative actions (McMurran & Theodosi, 2007; Olver, Stockdale, & Wormith, 2011). Sturgess, Woodhams, and Tonkin (2016) found those who dropped out perceived treatment as: boring, ineffective, repetitive, unnecessary, challenging, intrusive, stressful, patronising, and incompatible with their own goals. Similarly, Barnao et al. (2015) found that forensic service users experienced rehabilitation as lacking person-centredness and featuring relationships of varied quality. In order to address this issue, Barnao et al. (2016) used a case study to illustrate how a brief GLM programme may promote a more person-centred and holistic approach to treatment. Their findings suggested variation in outcomes associated with participants' readiness to change, their level of exposure to the GLM (i.e., the frequency and duration of sessions), and practitioners' adherence to and experience with the GLM. They found that those with greater exposure to the model through experienced practitioners had the greatest improvements in their perceptions of rehabilitation (Barnao et al., 2016). This finding supports the assertion that the integration of GLM throughout programmes and the training of practitioners in GLM concepts is crucial.

Other studies have found that both practitioners and participants prefer the GLM approach to treatment (Gannon et al., 2011; Harkins, Flak, Beech, & Woodhams, 2012), and that in terms of psychometric improvements and reducing risk they are able to perform as well as or in some cases better than standard relapse prevention approaches (Barnett, Mandeville-Norden, & Rakestrow 2014; Harkins et al., 2012). In addition, a number of small case studies have used the GLM with participants who had previously been difficult to engage in treatment, and found positive outcomes including programme completion, attending further treatment voluntarily, completing therapy tasks, constructing future plans, internalising risk-related knowledge and decreases in risky behaviours, and increases in prosocial goods attainment, wellbeing, and satisfaction in life (Lindsay, Ward, Morgan, & Wilson, 2007; Whitehead et al., 2007). Despite variation across participants and small sample sizes, these findings suggest that risk-focused treatment programmes may fail to engage some participants, and that GLM concepts (when used effectively) may produce shifts in perceptions of treatment and support a range of other positive outcomes.

7.3.4. Summary: Contributions of the GLM.

I have briefly summarised the key strengths of the GLM and some of the research providing support for its use in treatment. The most significant benefits identified in these studies include a focus on approach goals directed towards personally meaningful outcomes, and the engagement and commitment to change resulting from this motivational approach. In addition, due to their descriptive nature and attention to the perceptions of practitioners and participants, these studies highlight potential issues with implementing GLM treatment. For example, it is essential that participants understand the links between their GLP and risk reduction, and that practitioners are adequately trained in case formulation in order to support this process. These requirements are intertwined and dependent on an understanding of the goal-directed nature of persons and the range of capacities and resources which support goods attainment. This is where theories of agency can add depth to the GLM and support case formulation.

It is important to remember that the adoption of GLM concepts and agency theories in practice does not require abandonment of the RNR principles. Rather, programme designers and practitioners can integrate the most useful aspects of each model into treatment. As stated above, given the intended role of the GLM in augmenting the RNR model, if used properly it should be *at least equal* in its effectiveness, with the added benefit of being more motivating and potentially producing longer lasting changes. The aims of the justice system (i.e., reduced recidivism/harm) and the individuals who exist within it (i.e., PHG) are not incompatible. If programmes help individuals to develop the motivation, confidence, and resources to live a good life without causing harm, they will likely maintain an offence-free lifestyle in the long term. I suggest that the DRF-focused approach to reducing recidivism is largely failing, and it is time to adjust our tactic in light of the broader evidence concerning correctional interventions and human nature in general. Arguably, if a GLM approach to treatment appeals to those who use it (i.e., practitioners and participants), facilitates individualised and collaborative treatment, and can reduce some of the inherently negative aspects of interventions (e.g., externally imposed avoidant goals), then it is both a useful responsivity tool and an ethically important addition to correctional treatment. The question then is how to integrate these models with the reconceptualisation of DRF developed throughout this thesis, in order to take advantage of a wider evidence base and improve the way we design and deliver treatment.

7.4. Integrating the Risk-Need-Responsivity Model, Good Lives Model, and Agency Theories

I will now provide some suggestions for the integration of these rehabilitation frameworks and agency theories. Firstly, I would like to note that the use of DRF in informing risk level and intervention dosage (i.e., the risk principle) is not contested here. This is a concern for risk-prediction rather than explanation and intervention, and although issues of construct validity and other measurement problems persist (see chapter two), I will not address these further. However, as discussed earlier, the claim inherent within the need principle – that treatment should predominantly target *offence-related features* of individuals and their environments – is tentatively disputed, as is the assumption that we can identify and understand these fully through statistical analyses. I do not think that the features which correlate with offending (whether they are established causes or not) are necessarily the best or only targets for intervention. I agree with the general and specific responsivity principles broadly but suggest that they are limited in their scope if they fail to consider a wide range of available evidence. I will begin by briefly discussing the integration of the RNR and GLM before I outline the contribution of agency theories in supporting these rehabilitation frameworks.

As there are already numerous publications outlining the use of the GLM in treatment (e.g., Willis et al., 2013, Yates, Prescott & Ward, 2010), I will only summarise what the GLM *adds* to the RNR principles. As stated above, attention to addressing DRF or criminogenic needs (in a broad sense) remains a major goal of intervention, but this is explicitly integrated with the personally meaningful outcomes which make up a GLP. In terms of the risk principle, the GLM would not dispute the importance of risk assessment, but the scope of its application in treatment should be somewhat narrower. For example, informing the intensity or duration of treatment and pointing to aspects of persons lifestyles which may be risky – not directly informing the goals of treatment (this will be discussed further below). I suggest that DRF cannot guide treatment *on their own* – they are contained within treatment targets (i.e., approach goals supporting goods attainment). As discussed in the previous chapter, DRF are composite constructs and if they are to be used in treatment they should be 'stripped down' into their causal, contextual, and mental state facets. Integrating the GLM with the need principle involves the reconceptualisation of DRF as problems in the attainment of goods. In linking the GLM with the general responsivity principle, its core concepts (i.e., primary and secondary goods, approach goals) may be used alongside empirically supported techniques such as CBT and motivational interviewing. Adherence to the general responsivity principle requires that programme developers are responsive to the emerging evidence concerning what

works with this population, and this should include qualitative research into participants' experiences of different programmes. Finally, in terms of specific responsivity, the GLM can assist with motivation – participants need reasons to change, not just capacities to do so.

7.4.2. Implications of a focus on value-laden practices and agency capacities.

An important task when engaging in theoretical exploration is to consider how changes or additions to our understanding of key concepts might have an impact on people in the real world. I am concerned with how our understanding of DRF and PF influences the assessment, management, and treatment of individuals who have committed offences. I will now outline what the reconceptualisation of DRF (and PF) as aspects of agency (i.e., indicators of a more prosocial or antisocial orientation towards goal attainment) means in terms of practice. I will also consider the implications of identifying the causal processes underpinning agency, although significant work remains in this area. First, I will discuss the implications of an increased acknowledgement of values and the normative nature of DRF and PF. Then I will explore the potential implications of the PAM and the recommended search for mechanisms (i.e., the RCM) on case formulation and treatment. Finally, I will discuss some existing approaches to treatment which are consistent with this view of DRF and agency.

Values/norms/practices. Firstly, I suggest that theories of rehabilitation should be explicitly structured in terms of their normative, etiological, and practice components. This should make it easier to pinpoint the social and ethical commitments of different models and lessen the chances of conflating their empirical and normative claims. An example of this conflation is the positioning of DRF as causes of crime, when in fact they represent normative judgments about aspects of persons and their lifestyles which are 'antisocial' or 'crime-related'. Additional important value-based concerns include dealing with: 1) human rights duties and entitlements for individuals who have committed crimes (e.g., Ward & Birgden, 2007); 2) the problem of 'dual relationships' where practitioners often face conflicting ethical demands because they are committed to both mental health and criminal justice sets of norms (e.g., Adshead & Sarkar, 2009; Ward, 2013); 3) tensions created by the fact that people who harmed others have often been victims of crime themselves, something not readily dealt with in routine clinical practice (e.g., Ward & Moreton, 2008); 4) confusion between punishment and treatment and to what degree the former is evident in treatment programmes (e.g., Glaser, 2003); and 5) rehabilitation crucially depends on the concepts of DRF and PF, both of which are, in part, value-laden constructs (e.g., Ward & Fortune, 2016a). The existence of these ongoing controversies underlines the necessity of researchers and practitioners taking part in

ongoing normative dialogues with those directly and indirectly affected by crime. It is simply not enough to turn to empirical findings yielded by scientific inquiry.

Secondly, I propose that simply acknowledging the extent to which DRF and PF are value-laden is a step in the right direction, and that this is not currently evident in routine practice. One way for practitioners to prospectively recognise potential biases is to consider where they sit with respect to relevant values; for example, harm, consent, monogamy, respect, health, relatedness, sexual pleasure, success, and so on. Also relevant is where other practitioners who have supplied their own professional opinions sit; for example, colleagues may vary to the degree to which they share personal and professional standards (see Ward, 2013). Perhaps most important when working with individuals, is the ability to recognise any discrepancies between one's own values and normative standards, and those of the individual being assessed, treated, or managed by the justice system. These value clashes might manifest in therapeutic ruptures or misunderstandings depending on the task at hand.

Of concern for dynamic risk assessment is the subjectivity of ratings based upon practitioners' judgments concerning self-report and behavioural information (Cording et al., 2016). This is of equal relevance for treatment planning which is based upon assessments of risk/need. For example, despite the use of structured assessment manuals, there is room for disagreement in whether (and to what degree) a factor such as 'impulsivity' is present or relevant. The default assumption tends to be that risk factors are present, which results in homogeneity of risk profiles (i.e., all or most possible risk factors are identified) and little information about how to address the causes of offending (i.e., impairments in capacities and the contexts which they manifest within). A better understanding of the practices that DRF and PF refer to has the potential to facilitate theoretically informed explanations that apply to the individual and their values, rather than relying solely upon lists of correlates derived from aggregate data. A fuller picture of the influences upon these practices can guide interviews, help practitioners complete case formulations based upon shared and divergent values, and facilitate treatment which strengthens the capacity for practices which do not harm others.

Conceptualising DRF in terms of the goal-directed practices to which they refer, and the values and norms which form the context for these practices and their evaluation, may open fruitful avenues for new developments in the process and content of interventions. For example, rather than concluding that a participant needs to address their 'offence-supportive attitudes' (e.g., 'children are not harmed by sex'), practitioners can explore the values or goals underpinning these statements (e.g., inner peace, avoiding cognitive dissonance, pleasure), the norms that they violate (e.g., age of consent, the nature of harm), and the practices in which

they manifest in (e.g., intimacy seeking, offending, explaining/justifying). This is in line with Ward and Carter's (2019) suggestion discussed previously, that explanations are usefully directed towards motivational systems. These broader considerations provide additional information about the *context* within which harmful practices occur, the *function* of the particular feature in the individual's life, and eventually contribute to a more complete understanding of the *capacities* required for less harmful practices. This can help practitioners to ascertain whether and how a particular DRF or PF is likely to increase or decrease risk of reoffending for an individual, and in which contexts. The implications of this include more person-centred and holistic assessment, case conceptualisation, treatment, and management of individuals who have committed offences.

According to the view outlined in this thesis, treatment should focus on restoring the individual and their environment to a healthy (personally meaningful) level with respect to key areas of functioning, and the goal-directed practices which they underpin. Therefore, an important task for individualised treatment is to locate the source/s of impairments. For example, it may be that intimacy seeking strategies are normative, but that they have been directed towards an inappropriate target (i.e., a child) so they are labelled as 'grooming' rather than being seen as adaptive. In other cases, the individual might direct their attention towards appropriate targets generally, but in certain states or contexts they might act in ways that are incongruent with their beliefs and values. In the first example we see a clash in values and transgression of relationship norms, where the individual views children as viable partners – a perception not shared by the rest of society. In the second example we see impairments in self-regulation and control; the problem seems to be acting in ways that are incongruent with one's own values (as well as being against the law). The targets for treatment will be different in each case. For example, exploring the discrepancies between certain value systems, versus strengthening capacities required for effective self-regulation and coping with distress.

If practitioners fail to consider the preferences and perspectives of the people they work with, it is less likely that they will be able to be able to engage them in treatment to reduce the likelihood of further offending (Ward & Brown, 2004). In addition, by encouraging participants to identify their own values and sets of relevant norms, and to consider how they manifest in harmful or unhealthy practices, practitioners may engage participants in collaborative treatment. Participants can be positioned as the experts in their own lives, rather than being told that their goals are to reduce their 'offence-related attitudes and emotions' and 'poor coping strategies' (just like everyone else in the programme). It is expected that participants would experience this sort of treatment as more engaging, and that this would be reflected

within their motivation to complete treatment and remain offence free in the future (Ward et al., 2007).

In addition, this individualised approach will result in treatment that is truly culturally responsive, as it locates the person within various systems governed by norms. It is expected that culturally diverse groups (i.e., ethnicities, gangs, sub-cultures) are likely to endorse different practices and normative standards, and that such differences ought to be taken into account when a) developing risk assessment and psychological assessment protocols, and b) constructing treatment programmes. One advantage of conceptualising DRF and PF in terms of practices is the ability to locate offending behaviours within their cultural context, and to suggest ways of achieving valued outcomes without harming others or breaking the law. The process of correctional rehabilitation then becomes one of restoring or developing healthy functioning by strengthening capacities. It is a process of learning and building strengths, rather than simply punishment and risk-management. I have provided some general recommendations (see Table 3 below) for practice which are designed to expand upon the RNR principles.

Table 3

General Recommendations

Principle	Recommendation
Risk	Include dynamic risk assessment in judgments concerning treatment dosage and ground these within a theory of agency which can help specify the function of symptom-like problems (i.e., as risk increasing or decreasing).
Need	Case formulation should be holistic, focused upon a broad range of needs, and consider the function of offending (i.e., informed by agency theory). Orient treatment towards building agency capacities rather than just avoiding offending – both are important.
General Responsivity	Utilise research from a range of sources (including desistance literature) and include recent relevant findings. Remain critical of evidence and consider the limitations of research comparing eclectic treatment approaches.
Specific Responsivity	Ensure that those developing and delivering programmes have the appropriate knowledge and training to respond to individual needs. Avoid overly manualised programmes.

Agency capacities and case formulation. I will now offer some suggestions concerning how the PAM and RCM may influence the crucial task of case formulation. The reconceptualisation of DRF and PF advocated for throughout this thesis is essentially in line with the GLM concepts of internal (and external) resources and obstacles. They are symptomatic of capacities and contexts that either facilitate or obstruct agency, and thus direct persons towards or away from antisocial, harmful, or maladaptive outcomes. The PAM and RCM can add depth to the GLM by pointing to more specific areas where weakness (or strength) may be present. Even at this preliminary point in the reconceptualisation of DRF, practitioners may find these ideas useful to guide the assessment and treatment of individuals

who have committed offences. Simply relying on existing case formulation models or etiological theories that assume that DRF (as currently stated in the literature) are possible causes is likely to result in overly general, poorly integrated formulations (i.e., everyone looks the same or possible hypotheses are overlooked). My suggestion is that in the domain of intervention DRF should only be used to indicate *general problem areas* and be regarded as summaries of possible causes, contextual factors, behavioural, and mental state variables. By the processes of deconstruction, analysis, and reintegration we can put DRF to work in ways that are likely to give us a deeper understanding of why and how individuals act in ways that harm other people.

Case formulation is the analogue of theory construction in the practice domain; the creation of an explanatory model that accounts for the onset, interrelationships, and maintenance of problems associated with crime (Hart & Logan, 2011; Sturmey & McMurran, 2011; Ward & Beech, 2015). Forensic case formulation is used to guide the rehabilitation of those who have committed offences, and until recently has received relatively little attention in the literature (Sturmey & McMurran, 2011). It is an important task, which involves the gathering of risk-related data and construction of a plausible explanation (i.e., a hypothesis) for how and why particular factors cause and maintain offending. Typically, it is based upon the widely accepted 'propensity model' of risk, where a number of long-term vulnerabilities (i.e., DRF) interact with environmental triggers and opportunities to influence behaviour (Beech & Ward, 2004). Case formulation is generally guided by practitioners' training in cognitive-behavioural theory (CBT) and a general personality and cognitive social learning perspective of persons. There is an emphasis on faulty thinking, offence-supportive core beliefs, antisocial traits, and learned behaviour. In the sense that it links DRF with behaviour, case formulation aims to bridge the gap between prediction and explanation. It guides practice via the integration of empirical knowledge and theoretical understanding of DRF. The problem currently facing forensic and correctional practitioners is that existing case formulation approaches assume that DRF are theoretically coherent constructs (see Sturmey & McMurran, 2011), however, as argued throughout this thesis, they are not.

The purpose of case formulation is to identify factors which are relevant for the *individual,* and to use these to generate predictions for future behaviour and targets for intervention (Hart et al., 2011). Hart et al. (2011) provide a number of useful suggestions and evaluation criteria for forensic case formulation. Firstly, it is important that case formulations provide coherent narratives which are simple in that they avoid the incorporation of unnecessary data and assumptions, and diachronic in that they span the past, present, and future

rather than being fixated upon the past (Hart et al., 2011). Another important consideration is the explanatory breadth of a formulation, for example the extent to which it accounts for critical evidence and ties all relevant information together. It must be internally coherent, consistent, and contain assumptions which are compatible (i.e., do not contradict each other). Another criterion worth noting is a formulation's acceptability; whether it is useful and comprehensible for the individual for whom it was created (i.e., consistent with common-sense explanations rather than just complex scientific theory).

While forensic case formulation should be individualised in order to tailor treatment, in practice the hypotheses and goals formed often centre upon correlates of recidivism at the aggregate level (i.e., the need principle). This is particularly the case when assessments are undertaken by paraprofessionals (i.e., not qualified psychologists). This can result in case formulations and treatment plans which look remarkably similar across individuals; they lack explanatory depth and relevance because they rely upon descriptions of vague and composite constructs. Practitioners are often provided with a structured interview template containing a list of DRF deemed relevant for the individual if they feature in their offending. These are then translated into goals for treatment. For example, challenging and replacing attitudes concerning sex and violence, abstaining from or reducing substance use, improving emotion management, developing problem-solving skills, and so on. It is expected that these correlates cause (or at least contribute to) individual offending, and therefore that reducing or removing their influence will reduce the likelihood of reoffending. However, the empirical findings concerning the links between improvement on DRF and subsequent changes in offending behaviour are currently mixed (Duwe & Rocque, 2016; Serin et al., 2013), there are concerns about construct validity (Cording et al., 2016; Polaschek, 2016), and there is little understanding of the mechanisms underlying the change process.

Hart et al. (2011) suggest that forensic case-formulation should be *inferential*, involving abductive reasoning from data to phenomena and causal mechanisms, and *ampliative* in generating new knowledge rather than merely re-describing problems. Therefore, the development of mechanistic explanations is directly relevant to forensic practice, both in helping to locate the source (i.e., underlying cause) of impairments and also in understanding how these problems have developed over time. *Etiological* explanations focus on the process of an outcome being brought about, and typically refer to events or features which temporally precede an offence (i.e., potential causes). For example, loneliness and intoxication may lead to a sexual offence. *Compositional* explanations on the other hand refer to the structure of phenomena across multiple levels of analysis. For example, the functioning of the brain while

a person is experiencing loneliness constitutes the emotion, rather than causing it (perhaps it is partially caused by an external event and/or a previous thought). Case formulations typically provide an etiological explanation in the sense that they contain proposed distal and proximal influences upon behaviour, and predictions about the sorts of triggers and contexts which are likely to precede any future offences. I argue that these individual etiological explanations will be more useful if they are supplemented by compositional explanations concerning the phenomena implicated within these narratives. In other words, both are necessary for developing comprehensive explanations.

Theories of DRF and offending are therefore crucial within the formulation of individual cases, they guide the process of *inference* from *descriptions* of thoughts and behaviours to *explanations* of their possible causes and targets for intervention. As they are partially informed by narrative data (i.e., the person's account of the offence), these explanations can be thought of as comprising first, second, and third person perspectives. While the offence itself is usually not able to be observed by the practitioner or researcher formulating the explanation, they may rely upon the perspectives of the individuals involved (i.e., perpetrator and victim), others present (e.g., witnesses, law enforcement), and others involved in previous assessments (e.g., Corrections/justice staff). In addition, they rely upon empirical data which has identified common areas of difficulty for those who have offended (i.e., DRF), and opinions of their relevance to the individual. These are inferred from the perspectives of the individual themselves, those who know them well, or practitioners who have worked with them previously. Ideally explanations will combine evidence from numerous high-quality sources in order to form a comprehensive and accurate formulation that is guided by a sound theory (or theories) of human functioning (Hart et al., 2011). This would also contribute to a formulation's reliability; the likelihood that others provided with the same information would come to the same conclusions (Hart et al., 2011).

Similarly, it is important to consider the levels of explanation which are most useful for the task at hand. For example, the mechanisms underpinning offending are comprised of genetic, neurobiological, psychological, social/cultural, and ecological systems (Weerasekera, 1996). Persons are comprised of hierarchical systems, with unique properties emerging at each level. This means that explanations pitched at just one level and ignoring others will be impoverished; all will form a piece of the explanatory picture. For example, in explaining a certain norm-violation, it would be a mistake to overlook the importance of the physical and social environment (e.g., laws, norms, opportunities, etc.), the psychological features (e.g., emotion, cognition), or the biological and neurological influences upon cognition and

behaviour. While one level may not be reduced to another (e.g., anger cannot be fully explained by physiological arousal) a certain level of analysis could be more suited to a particular task. Certain levels are influenced by others in a process of upward (e.g., biological influences on mood) and downward (e.g., environmental stressors influencing thoughts) causation, reflecting both horizontal and vertical interactions (i.e., between mechanisms and across levels of analysis).

I suggest that *first person*, *intentional* (psychological level) explanations are privileged for the purposes of case formulation, and these should focus on descriptions of behaviour – what people do or fail to do, and why. While biological and social levels are important, we ultimately need to understand *why the person acted* in ways that harmed others. This means that useful explanations of serious norm violations and their potential causes must refer to mechanisms and offer an explanation which differentiates adaptive or optimum functioning from dysfunction (Ward & Fortune, 2016a). Further, these explanations must be able to account for the interaction of mechanisms (within various DRF categories) which result in norm violations, it is not enough to specify the various mechanism components without a satisfactory account of their causal trajectories. In addition, these explanations will contain both factual and normative components. Factual or causal explanations depict the parts and processes of the mechanism/s comprising a phenomenon, the normative component spells out whether or not it is performing according to some set of standards (i.e., functioning spans a continuum). A research framework such as the RCM has the potential to both guide the development of compositional explanations within individual case formulations (i.e., via local models), and also to understand how various mechanisms are linked with each other and behaviour through their integration within a model of agency (i.e., the PAM).

In order to make these implications more concrete, I will now make some preliminary suggestions concerning how the different phases or tasks of practice can be responsive to the problems with DRF and the reconceptualisation developed throughout this thesis. First, as they are based on DRF, previous risk assessments may provide guidance on problem areas to be explored. However, it is important not to jump to conclusions about the causes of behaviour, or to rely solely on DRF identified through risk assessments to guide formulations. Rather, practitioners can use these ratings as one source of information (i.e., as potential barriers to prosocial agency), and later check whether the DRF identified via risk assessment are likely to be linked with or addressed within their case formulation. The most important source of information for constructing the case formulation will be the individual's account of their behaviour, including the motivations and values underpinning it and the normative context in

which it occurred. In line with the GLM, the PAM (and future local theories) can be used to formulate a good life plan (GPL). The key tasks are to determine the individual's good life conception (i.e., priorities for living), view these in light of their offending (e.g., direct or indirect routes, problems evident), and then develop goals which centre upon the internal and external resources required to live a good life without hurting others (i.e., a GLP). For examples of key areas of inquiry see Table 4.

Table 4

Components of the Predictive Agency Model, Good Life Plan, and Considerations for Inquiry

PAM	Good Life Plan	Additional Example Questions	Potential Areas of Vulnerability
General models	What does a good life look like? What is important? Are all needs covered? What resources are needed?	What general models were salient or available at the time? Are these normative? Adaptive? Harmful?	Intra- and inter-personal systems Cognitive systems – e.g., attributions of meaning and causes of events, biased/inflexible general models General models which are harmful or obstruct needs attainment
Emotional lens	Why is this important? What emotions are attached to outcomes?	How is emotion generally? (i.e., positive/negative affect, slowly/rapidly changing, etc.)	Positive and negative affective systems – e.g., threat detection, reward seeking, sense-making
Context	Does the context support meeting needs? What external resources are required?	What was the offending context? How did the individual make sense of the situation? (i.e., local models)	Contexts which reward crime Contexts which obstruct or do not offer opportunities to meet needs
Motivation	In what situations are certain needs salient?	What value(s) or needs were triggered in the situation?	Scope of GLP Prioritising some PHG over others
Planning	How can needs be met? (e.g., secondary goods) Are there multiple strategies? What influences action selection?	Was this more implicit/explicit? What were the options? What was the expected reward? What was expected by others? Was it achievable? Why? Why was this option chosen?	Self-regulation systems Action selection and control Inhibition Creativity and flexibility
Action/ Feedback	What are the positive and negative consequences?	What happened? How did the individual feel? Was it rewarding? How did others respond?	Cognitive systems – e.g., attributions of the meanings and causes of events
Reflection	Were there problems? Is this behaviour in line with values? Are there other ways to meet these needs?	What did the individual learn? Did the outcome match their expectations? Has anything changed?	Learning capacities Knowledge integration General model revision

Initially, the practitioner should conduct an exploratory, collaborative, and semi-structured interview to draw out the different components of the GLP. This may be guided by the PAM or a similar model of behaviour. For example, the practitioner will consider the person's background, early experiences, and how these have shaped identity, interpersonal style, and world view. They will explore the individual's priorities, how they conceptualise a good life and how they have in the past strived to reach this, which may relate in different ways to offending. Thus, this initial stage introduces a dual focus on individual well-being and risk. The practitioner will consider emotional functioning, the role of emotion in motivating

behaviour within certain contexts and explore local models or situation specific representations of the offending context. This will lead into a discussion of the planning phase, considering the expectations which (implicitly or explicitly) influenced the choice of secondary goods. The individual can then describe the behaviour and its immediate consequences, including responses from others, and their initial thoughts and emotions. This will help to understand how the behaviour fits with their implicit GLP, for example whether and how the behaviour was rewarding (i.e., was it successful in meeting needs?), and if it was congruent with expectations. These sources of internal and external feedback can then be discussed, and their impact on general models explored.

When it comes to setting goals and delivering interventions, practitioners can use comprehensive case formulations to develop goals based on the specific causes identified (i.e., hypothesised areas of vulnerability) and the individual's GLP. They can locate the potential sources of problems and then draw from available compositional (i.e., local) models to increase specificity and understanding. This step may involve additional levels of explanation – including social, contextual and behavioural, as well as sub-personal mechanisms (which levels are relevant will depend on the problem and task at hand). The strategies for developing strengths or resources will vary. For example, vulnerabilities in the planning stage may be addressed by increasing options for action selection, and problems with general models may be addressed by shifting these to more healthy, adaptive, or prosocial alternatives. According to the PAM, behaviour will change when expectations change, and expectations change via feedback signals. Therefore, rehabilitation should aim to provide learning opportunities or situations where error signals are experienced, so as to alter expectations for future scenarios. For this to happen individuals need to possess the capacities required to learn from these situations, and so this process (i.e., feedback and reflection) may need to be guided and supported by therapists. It is acknowledged that individuals participating in correctional programmes often display problems with self-regulation and planning. However, the majority appear to benefit from training in problem solving and cognitive training, and therefore clearly possess the basic components of agency. I suggest that with support capacities can be built which allow the individual to explicitly identify and implement alternative strategies (i.e., new secondary goods), and that with practice (and reward) these may become part of their implicit GLP. Thus, this approach advocates for a focus on building capacities and resources (internal and external), including shifting the way the individual represents themselves, others, and the world, in order to support needs attainment and their conception of a good life.

Compatible treatment approaches. There are existing treatment approaches which are compatible with the PAM and its view of feedback and reflection as crucial in changing, updating, or shifting the general models which guide behaviour. Correctional treatment generally takes the form of group-based cognitive-behavioural interventions, with a focus on patterns of past offending behaviour (e.g., developing and analysing offence maps or offence chains) and safety planning for future behaviour (i.e., relapse prevention approaches). In this sense, these interventions are already somewhat future-focused, but arguably they could be more so. If we view offending as future-focused rather than driven by past behaviour, the emphasis of treatment would be on future outcomes which meet needs without offending (e.g., approach goals within a GLP). Thus, interventions which are consistent with the GLM and its view of persons as goal-directed agents are clearly compatible with the PAM, but so are other widely used techniques.

Seligman et al. (2016) suggest that CBT can be adapted to be more future-focused, in line with the prospection approach. For example, by exploring 'if-then' statements (i.e., causal attributions) or generating alternative simulations of the future and evaluating these. According to Seligman et al. (2013), there are three key aspects of prospection, the complexity and flexibility of plans, length of the time horizon, and the accuracy of expectations. It is important that individuals are capable of developing multi-faceted options for action (e.g., goals, sub-goals), and that these are flexible to environmental contingencies. In addition, the breadth of options available to a person is important and this relies on the individual's creative capacity as well as past experiences. It is also important that the individual can extend the time horizon of their plans and foreseeable consequences (e.g., long-term thinking). Finally, it is important that expectations are accurate and that individuals are capable of second order desires (e.g., wanting to be motivated towards things).

Key therapist tasks according to this approach include challenging or replacing maladaptive general models, increasing alternative options for planning, strengthening long-term thinking, learning through experience, planning for a better future, increasing incentives for change, and building meaning or purpose (Seligman et al., 2013). The ability to engage in these tasks relies upon a number of capacities, including (but not limited to) creativity and flexibility (e.g., generating alternatives), perspective-taking and other aspects of empathy (e.g., emotion recognition and regulation), self-regulation and delaying gratification, counterfactual and critical thinking, and hope and optimism for the future. There are existing interventions which aim to develop these skills, and I suggest that these will be more effective when underpinned by a comprehensive understanding of human nature and functioning. In addition,

an approach to treatment which strengthens an individual's sense of agency or autonomy will lead to a more internal locus of control and experiencing more control over behaviour and its outcomes. According to Seligman et al. (2013), when people feel like they have more 'free will' they report engaging in more long-term thinking and conscious deliberation, act in ways that are consistent with their values, and experience more positive outcomes.

An example of a treatment task in line with this approach is the use of 'behavioural experiments' (Gannon, 2016) within cognitive-behavioural interventions. Gannon (2016) suggests that challenging problematic (i.e., norm violating or maladaptive) attitudes and beliefs through experiential learning will be more effective than conversational or educational approaches alone. The idea is that through actually experiencing consequences which violate expectations, individuals may shift their representations of the self, others, and the world. For example, where others do not respond as expected or behaviour is not experienced as rewarding. Or alternatively, if an individual experiences success in meeting their needs in alternative ways then these may become a part of general models and used in the planning process in future situations. In a sense, this is a more purposive version of the agency process; when individuals engage in prospection, they are essentially generating testable predictions and then assessing their accuracy (e.g., through behaviour, feedback, and reflection). Therapists can guide this process by helping individuals expand on their behavioural options during the planning phase, and then observing and making sense of the results of these experiments, thereby scaffolding the process of updating general models.

In addition, the interactional nature of the PAM is consistent with suggestions that situational and environmental factors be included within intervention and prevention efforts (Smallbone & Cale, 2016). It is important that environments support internal changes through rewarding prosocial behaviour (e.g., incentivising change), and that systemic factors do not get in the way of desistance by obstructing goods attainment (Hannah-Moffat, 2016). Persons need opportunities and reasons to want to change as well as the capacities to do so (i.e., external as well as internal resources). Incentives can be deliberately or professionally provided (e.g., lessening restrictions, work opportunities) and/or naturally occur in environments that reward desistance more than crime. Thus, interventions should explicitly address the reintegration process and develop expectations for this experience which support desistance. For example, encouraging hope and optimism through positive simulations of the future within a GLP, and strengthening skills for creating or seeking out environments which support this.

The compatibility of the PAM with these existing approaches is likely a strength, considering the empirical support for the efficacy of CBT in behavioural change generally and

with offending populations (Andrews & Bonta, 2010; Bonta & Andrews, 2017). However, the PAM expands upon and adds depth to the cognitive-behavioural model by spelling out how enduring general models (e.g., 'core beliefs' in CBT) are utilised throughout the agency process, as well as adding to our understanding of the relationship between cognition and emotion. For example, the role of emotion as sense-making and action-guiding is elaborated upon, consistent with emerging research. The PAM is also able to better incorporate the role of motivation and planning within the agency process, and in doing so places autonomy and responsibility for change within the person – as opposed to avoiding or reacting to external sources of risk (i.e., triggers, high risk situations in CBT). In this sense the PAM arguably offers a more sophisticated and comprehensive understanding of cognition and emotion, and the reasons for human behaviour.

7.4.3. Summary: An Integrated Approach to Treatment.

In summary, it makes little sense to construct case formulation and intervention plans without a reasonable understanding of what the possible causes of crime and related phenomena are. Relying on DRF as treatment targets is a mistake as they do not reliably identify underlying causes at all; they are in effect *summary labels* for possible causes, contextual features, behaviours, and mental state variables. Strictly speaking, DRF do not exist for the purposes of treatment, and there is little point targeting symptom-like summaries and assuming this will alter the mechanisms generating them. However, from a pragmatic viewpoint the PAM and RCM can play valuable roles in structuring clinical inquiry and, in conjunction with knowledge of etiological theories, risk assessment, and classification models, can assist practitioners to arrive at a working explanation of an individual's crime-related problems. It can bridge the gap between risk assessment and intervention and ensure that practitioners carefully consider the explanatory possibilities offered by DRF and avoid the trap of assuming they directly pick out causal factors. An advantage of structuring assessment and subsequent treatment in this way is that it confers a degree of epistemic scepticism on practitioners' conceptualisations of clients and reminds them that they critically depend on theoretical and methodological assumptions.

This chapter has explored the implications of this approach for case formulation and treatment in a forensic setting. It loosely sketched out an approach to treatment based on the GLM, the PAM, and future integrated theories (e.g., the RCM). This approach is intended to be complementary to the RNR model, but also to add depth concerning the mechanisms underpinning criminal behaviour (i.e., the need principle) and integrate these within a

comprehensive model of human behaviour which can guide individualised (i.e., responsive) treatment. In this approach DRF are viewed as predictive devices which may indicate areas for further inquiry, and they are likely to demonstrate changes following effective treatment. However, treatment is more broadly focused on strengthening the internal and external resources required to meet needs in prosocial and healthy ways.

7.5. Conclusions: Implications for Practice

The question of what constitutes effective treatment for persons who have committed offences remains, and the evidence used to answer it must come from a range of sources and methods of inquiry. The existing research investigating the use of the RNR and GLM is promising, but it is by no means conclusive, and there are problems with research methods and the implementation of rehabilitation models in practice. For example, many GLM-consistent programmes fail to use the model as intended (i.e., to guide the entirety of the programme), and the RNR model is routinely reduced to the three core principles and implemented variably internationally (Polaschek, 2012). Future application and evaluation of these models (and agency theories) must integrate them as intended and should evaluate various aspects of treatment and a range of outcomes of interest using multiple methods.

It is apparent that "there is much that is unknown about what is effective in reducing sexual and violent recidivism, and it is possible that the content included in the program model used is not effectively targeting the appropriate issues or risk areas" (Grady et al., 2017, p. 259). The variable and modest effects of interventions suggest that progress is necessary, and we must continue to learn from evidence and conceptual issues in other related disciplines, as well as empirical and theoretical developments in forensic psychology. I suggest that research aiming to advance rehabilitation should not rely solely upon lists of correlates and statistical relationships; the variables involved in the process of change are simply too complex. The RNR model has achieved substantial improvements over the years and is based on an impressive amount of evidence, but it is not the final word on what works to reduce recidivism (Newsome & Cullen, 2017; Polaschek, 2012). The RNR principles are based on much of the evidence gathered thus far, but this should encourage critical attention and further development, rather than acceptance of small effect sizes. This is especially important considering the consequences of treatment failure and the likelihood of unmotivated or disengaged individuals giving up on the possibility of change.

Chapter Eight: Conclusions and Future Directions

Dynamic Risk Factors (DRF) are the most utilised concepts in the field of forensic and correctional practice. They provide the theoretical foundation of the RNR model and are the identified foci of correctional treatment programmes. This thesis has built upon recent critical attention towards these constructs and questioned widespread assumptions concerning their relationship with risk and recidivism. I have provided a preliminary model for viewing DRF, protective factors (PF), and offending as features of norm- and law-violating agency processes, and made suggestions concerning what this means for practice moving forward. I will now summarise the contributions of each chapter of this thesis and then evaluate this approach to explaining DRF and offending. I will conclude with some suggestions for future research.

In chapter one I provided an overview of the conceptualisation of DRF and PF, and their dual roles in practice, both as predictors of recidivism and targets for treatment. I argued that their transportation from the realm of prediction to treatment relies upon their status as causes of offending. In chapter two I explored their causal nature through considering the empirical evidence concerning DRF, change, and recidivism. This chapter found mixed support for the relationship between treatment, DRF change, and recidivism. Unfortunately, this evidence was undermined by methodological weaknesses and the fact that in order "to test hypotheses about the causes of sexual offending, we must not only demonstrate a causal relation between variables but must also have a clear and concise conceptualization of the construct of interest and be able to accurately measure it" (Nunes et al., 2019, p. 231). We currently do not have clear and universal definitions of DRF due to their composite nature.

In chapter three I discussed conceptual issues with the DRF construct and problems with existing theories relying upon DRF to explain offending. Due to these issues, in chapter four I concluded that in their current state DRF are unable to function as causes of offending and compared them against a list of causal criteria. I demonstrated that while DRF may predict and precede offending, they do not display the expected relationships with recidivism (i.e., following changes through treatment) and they are lacking in coherence and specificity. This means that they are unable to pinpoint causal mechanisms as they are normatively defined composite categories. They contain a range of things (i.e., mental states, contexts, behaviours) and are only considered DRF due to their co-occurrence with crime (i.e., many of them are also present in non-offending populations). Due to these issues, I argued in chapter four for an initial shift in focus from explaining DRF and crime, to explaining goal-directed practices (some of which are crimes).

In chapter five I presented the *predictive agency model* (PAM), which depicts the agency process and the capacities involved in goal-directed practices. While the PAM is a step in the right direction in terms of shifting the focus from correlates to goal-directed behaviour and the capacities which support human functioning, it was necessarily shallow and did not specify the mechanisms underpinning the agency process. In order to address this issue and add depth to the model, in chapter six I developed an approach to research building upon several recent suggestions for theory development. The resulting framework, the risk-causality-method (RCM), suggests breaking DRF down, extracting their possible causal components, and then developing local models of these causal processes across multiple levels of analysis. The result is a set of local models which can be integrated within a model of behaviour, such as the PAM. In chapter seven I discussed some implications of the practices (i.e., values and norms), PAM, and RCM approaches for forensic practice. I proposed that by integrating the most promising aspects of the RNR, GLM, and agency theories, we can improve upon existing approaches to treatment and more readily meet the aims of the criminal justice system.

8.1. Overall Evaluation

I will now offer an initial evaluation of the approach to DRF and PF developed throughout this thesis. In brief, this reconceptualisation views DRF and PF as predictive constructs which are symptomatic of underlying capacities (or lack thereof) and environments. They are summary categories for areas of vulnerability or strength, things which either support or obstruct prosocial agency. For this reason, they tend to co-exist with patterns of behaviour (including but not limited to offending and desistance), but they are not causes in any straightforward sense. They are potentially caused by a range of mechanisms which span multiple levels and exist within and outside of individuals. I will draw from a set of theory evaluation criteria, some of which were discussed earlier in relation to existing theories of offending. A preliminary evaluation of the PAM was offered in chapter five and the points below will elaborate on this evaluation while including the potential contribution of research in line with the RCM. It is important to remember that theory evaluation is a comparative task. We are not searching for 'one true theory' (Ward, 2019), but rather comparing this approach with existing theories of DRF and offending. At times I will refer to this overall approach as "agency theories", reflecting that it includes both the PAM and future theories developed in line with the RCM which add depth to this model. I will discuss the relative strengths and weaknesses, and also consider the extent to which agency theories can address the issues with DRF and PF outlined in earlier chapters.

Firstly, I suggest that the PAM demonstrates strong *internal coherence* and *simplicity*; it avoids contradictions and is relatively easy to understand. This is due to its general nature and the fact that it is pitched mainly at the phenomenological level of explanation. It is possible that there may be some confusion around some of the processes involved once depth is added via the RCM. The PAM has arguably added some *explanatory depth* on its own by pointing towards or sketching out potential mechanisms underpinning agency. For example, the critical role of emotion in the formation and retrieval of general models, as well as in the planning process alongside different kinds of expectations. This detail adds depth to the general idea of costs and benefits or "definitions favourable to crime" (Andrews & Bonta, 2010). However, further depth will need to be added by future research guided by the RCM. The suggestion to deconstruct DRF and analyse possible causal processes is in line with current approaches to explaining psychopathology (i.e., the RDoC), thereby demonstrating *analogy* with comparable research domains and problems.

The *explanatory breadth* or *scope* of a theory refers to its ability to explain a range of phenomena. I suggest that agency theories can explain a wider range of behaviour due to their broad focus on human functioning and goal-directed behaviour, rather than narrowing the target of explanation to sexual offending and DRF. Thus, in addition to offending, researchers or practitioners may apply agency theories to the process of desistance and other relevant behaviours (e.g., substance abuse, denial). This broad focus means that not only can agency theories be applied to a wide range of behaviours, but they can also be highly individualised and account for the influence of culture on individuals and their environments (e.g., through norms and the kinds of general models developed). As discussed earlier, the PAM has already generated and informed further research into cultural differences in DRF (Schmidt, Heffernan, & Ward, 2020) and the development of a cultural-ecological model of agency (Strauss-Hughes, Heffernan, & Ward, 2019), suggesting that it has *heuristic value* or *fertility*. It is hoped that the RCM will encourage and guide future research into the mechanisms underpinning DRF categories.

External consistency and *unifying power* refer to the extent to which a theory is compatible with existing knowledge and its ability to draw together different theories and evidence. I suggest that agency theories can do this relative to other theories of offending, as they are consistent with (but add depth to) approaches such as the psychology of criminal conduct (and RNR), good lives model, and cognitive-behavioural therapy (see the previous chapter). Due to its origins in the literature on prospection (e.g., Seligman et al., 2013, 2016) the PAM is also consistent with the impressive body of evidence which this work draws from.

Therefore, while the extent to which this approach is empirically supported is yet to be determined (e.g., via testable predictions of the RCM such as those suggested in chapter six), it can be considered consistent with current neuroscientific evidence concerning the roles of cognition and emotion in human action. Agency theories are also consistent with research and theory which depict the relationship between individual dispositions and situational factors (e.g., Smallbone & Cale, 2016), as they acknowledge the role of the context in both triggering motivation and shaping the development of general and local models used for planning. In addition, this approach is built upon previous theories of agency, and is consistent with earlier theories of sexual offending without being focused so narrowly on the offence and DRF.

Agency theories are also compatible with existing empirical research into the correlates of offending and desistance, while making sense of some inconsistencies of concern to researchers. For example, they can account for research showing differences between crime causing and crime reducing factors (e.g., Kroner et al., 2017) through a focus on the agency process across time (i.e., the pathway into and out of crime differ due to experiences and shifts in general models). Agency theories can bridge the gap between the predictors of crime and the process of desistance to some extent, by framing both as aspects of goal-directed behaviour; differences observed are largely an artefact of research priorities and methods. For example, DRF are correlates of offending identified at the aggregate level for the purpose of predicting individual risk or allocating individuals to groups for intervention (i.e., the risk principle). Desistance factors (and some PF) are derived from research looking at life events and narratives of those who have developed a more prosocial orientation (i.e., cognitive transformation) following offending. For these reasons they identify different features, but all are aspects of goal-directed human functioning (i.e., predictive agency) within different arenas of life (e.g., relationships, leisure, employment, etc.). Further, agency theories can help us understand the mixed and inconsistent findings outlined in chapter two, as well as the differences found in the accuracy of risk assessment across cultures and genders. The reason for inconsistencies in DRF measurement is that they are composite categories reflecting agency capacities and their manifestation in goal-directed practices, they lack specificity, and they are normatively defined. Therefore, they could not possibly demonstrate a predictable (i.e., linear) relationship with behaviours and they are not valid predictors outside of the normative cultural context within which correlates were identified.

Finally, the PAM and RCM were developed to address the problems with DRF discussed earlier in this thesis. For example, they overcome the issues of incoherence, composite constructs, and the grain problem by explicitly deconstructing DRF into their

component parts. These problems are further addressed by prioritising the first-person or phenomenological level of explanation, with other levels are to be addressed through research in line with the RCM. In addition, agency theories can begin to explain why sexual offending occurs as opposed to other behaviours (discussed above), through the development and strengthening of general and local models (i.e., possible actions), and specificity will be increased once depth is added by the RCM. They can address normative concerns by locating these capacities within a dynamic interactional model (i.e., depicting motivations and environmental constraints/opportunities) which can account for different kinds of practices, irrespective of whether or not they are legal. As discussed above, the dual status of DRF and PF is addressed by viewing both as aspects of agency, which can be directed at either prosocial or antisocial outcomes.

8.2. Future Directions

The main suggestion I will put forward for future research is that researchers utilise a methodological framework such as the RCM to guide the decomposition and exploration of crime-related problems (i.e., DRF). The exploration of possible causal factors using the RCM provides a group-level or *nomothetic* theory-building approach, which can then be used to help practitioners construct individual case formulations. In turn, analyses of specific case formulations may result in the discovery of additional possible causal factors, and interventions may provide further tests of the validity of causal assumptions – an *ideographic* theory-building approach. In addition, the use of agency theories within case formulation and treatment should eventually be expanded upon through the development of clear guidelines for practice. This will ensure that these concepts are used consistently and as intended.

Relatedly, it is necessary to investigate the potential utility of the PAM and future local theories in informing correctional treatment and management of individuals within the community. This is a particularly important task because (once expanded upon) the PAM has the potential to improve practitioners' and participants' understanding of the causes of offending, enhance the therapeutic relationship, and ultimately to promote behavioural change in a more meaningful and enduring way. My hope is that this approach will eventually offer a view of criminal behaviour that capitalises on decades of research into the nature of human functioning, rather than seeing those who engage in criminal behaviour as fundamentally different to the rest of society. In my opinion, this shift is overdue, and is one way forward in response to the theoretical dead end the field has recently encountered.

Examples of testable predictions derived from the PAM (and future theories developed according to the RCM) include the following: 1) once broken down into *functional units,* DRF and PF should shift in ways that reflect progress or no change across the offence and treatment process; 2) the manipulation of emotional states should result in shifts in cognitive factors and behaviours in the predicted directions; away or towards increased risky decisions (e.g., computer simulation studies); 3) the use of the PAM is expected to improve the quality of clinicians' case formulations compared to those relying on the standard characterisations of DRF and PF; and 4) because individuals are hypothesised to use local (i.e., situation specific) models to predict outcomes and guide action it is expected that features of environments will strongly influence decision-making. It is also possible that measures of agency from other domains (e.g., poverty and empowerment, see Ibrahim & Alkire, 2007) can inform the investigation of predictive agency (e.g., control, efficacy, and autonomy) in relation to the experiences of forensic populations. Thus, future research requires a multi-disciplinary approach, and the integration of a range of explanations with various targets and levels of analysis.

As discussed in earlier chapters, Sullivan (2019) suggests that interfield cooperation is required in order to move the field towards a more comprehensive understanding of crime and related constructs. This involves researchers clearly defining and operationalising constructs, and communicating across fields in order to add depth to our understanding. In order to integrate various levels of explanation, researchers must spell out the links between levels – for example, specifying the relationship between biological and psychological mechanisms and how they might give rise to behaviours. This is no easy task; researchers have long struggled to elucidate the relationship between these levels of analysis. Thomas and Sharp (2019) state:

> ... this lack of understanding regarding how to relate psychology to biology is a major reason for the dearth of causal explanations of psychopathology. Although this issue remains unresolved, field-wide attention has been primarily focused on whether findings are reproducible. Demonstrating reproducibility, however, does not suffice for building strong causal theories. A highly stable correlation between an intervention and an outcome may do little in explaining why the treatment works, which is necessary to further advance for whom and under what conditions it works best. In general, causal theories are vital to improving predictions regarding how psychological processes unfold under various conditions and

creating more precise interventions on psychological functions that go awry across various forms of psychopathology. (p. 2)

These issues highlighted in the field of psychopathology are analogous to the lack of causal explanations in forensic psychology. This further necessitates the interfield co-ordination to provide depth to explanations of offending, as proposed by Sullivan (2019), and highlights the complexity of this research programme.

Relatedly, recent attempts have been made to elaborate upon the theoretical basis of the RNR. For example, enhancing the model through integrating biological and social explanations (Carter & Mann, 2016; Newsome & Cullen, 2017). I suggest that this research is a step in the right direction when it comes to developing multi-level explanations of crime, and that it may offer additional areas for inquiry in research programmes such as the RCM. In addition, it has been suggested that of the core RNR principles, responsivity is "the least developed of the three. It is theoretically unsophisticated: a catch-all category" (Polaschek, 2012, p. 8). Thus, I believe it is crucial that future research prioritise the elaboration of this principle, as it encompasses important issues such as motivation and cultural needs. A more comprehensive understanding of the perspectives and priorities of potential participants can help practitioners overcome problems such as high rates of drop-out and poor engagement. In addition, while criminogenic targets *can be* framed positively (Polaschek, 2012), they are still an externally imposed list of treatment goals that are not routinely linked with personally meaningful outcomes other than the potential consequences of reoffending (i.e., future prison sentences). It would be useful to look at the impact of language and orientation of treatment goals on participants' expectations of treatment, subsequent engagement, and other meaningful outcomes (i.e., PHG attainment). Questions which would enhance our understanding of what works include: why do people change? Which aspects of treatment are most useful to individuals who want to change? What sorts of changes are most important in reducing risk? And which internal and external resources best support the process of desistance from offending? The answers to these questions can help us to develop more balanced approaches to treatment, where the process of change is considered as important as the focus on past behaviour and likelihood of future harms.

In addition, I suggest that desistance research may be able to provide answers to some of these questions. Briefly, desistance may be defined as the on-going process of shifting from active offending to reductions in and eventually the cessation of offending altogether. Features which have been observed to precede or facilitate this process include marriage, employment,

and the development of a new prosocial identity (McAlinden, Farmer, & Maruna, 2017). The desistance research also suggests that the development of a coherent narrative explaining why the offence occurred and why it will not happen again can help an individual to make sense of their past and take control of their future (Maruna, 2011). Agency and autonomy are necessary for individuals to take control of their future and develop the internal and external resources necessary to make shifts in identity and attain needs in prosocial ways. The resources associated with desistance (e.g., relationships, employment, self-control, good problem-solving skills) are often referred to as protective factors (PF), and we currently lack an adequate understanding of how they function to reduce risk (Ward, 2017).

I suggest that the GLM and the predictive agency model (PAM) can help us understand the relationship between risk and protective factors. The concepts of PHG and agency capacities can enhance our understanding of how PF or desistance events such as employment and relationships can reduce risk, and why their influence may vary across individuals and situations. For example, employment could be thought of as a secondary good which meets a range of needs, including (but not limited to) excellence in work and a sense of achievement, agency, life (i.e., financial resources), and inner peace (i.e., freedom from stress). Employment may also be a result of the development or strengthening of agency capacities such as conscientiousness, self-control, communication skills, etc. Employment which does not reflect these strengths or meet these PHG is less likely to support desistance and may lead to the use of other secondary goods (e.g., substance abuse, theft, dishonesty). The use of a GLP in treatment can be linked with possible futures within the desistance literature. For example, individuals must be able to conceive of a personally meaningful and attainable future and construct a new identity – "the self is continually being projected into the future" (Farrall, 2005, p. 369). This is a key commitment of the predictive agency perspective (i.e., the mind as a predictive engine), and in this sense the PAM can be used equally to explain offending and desistance. It can help us understand the reasons for change and how this process might occur, for example, what needs are motivating change and what resources does a person require.

I suggest that when it comes to informing interventions and explaining prosocial change, neither DRF nor PF on their own can provide the full picture – both offer one source of information concerning an individual's ability to meet their needs in prosocial ways. What is needed is a model that can account for why we observe DRF and PF alongside various behavioural patterns or trajectories of crime and desistance. I suggest that agentic perspectives such as the PAM can do this through their focus on goal-directed behaviour generally, rather than value-laden outcomes. Thus, research into the desistance process can provide useful

targets for explanation and intervention and it may also provide useful starting points for research into the mechanisms underpinning desistance and thus should eventually be included alongside DRF as instigators of inquiry in the RCM.

Finally, I suggest that future research should aim to elaborate upon why some individuals choose sexual offending (or violence) as a means to meet their needs, while others do not. This may be done by elaborating upon the development and maintenance of general and specific models and their role in action selection. For example, how are these formed and how can they change? Do they represent enduring models (i.e., sexual offending is always an option) or temporary impairments? It is a general weakness of previous theories that they fail to specify why some people use sexual offending to meet needs (i.e., for intimacy, pleasure, control), while others choose other more or less adaptive behaviours. It is likely that this is due to past experiences and expectations available to draw from during planning for action, and that these differ across individuals. These processes can be investigated via in-depth interviews with individuals who have committed sexual offences (or other behaviours) which may help researchers to understand the development and revision of these models over time. For example, by asking questions about early sexual experiences and the frequency of thoughts about sexual offending (i.e., the availability of general models supporting sexual offending). In addition, researchers could look at the differences between the planning abilities (i.e., generation and selection of options) for individuals who have sexually offended and individuals who are motivated for sexual offending against children but refrain from acting on these preferences (e.g., non-offending individuals with paedophilic interests). This would help researchers to better understand the relationship between sexual preferences, motivation, and planning.

8.3. Final Conclusions

The reconceptualisation of DRF and PF developed throughout this thesis has the potential to take the field of forensic psychology in a new and exciting direction and prevent us from encountering an otherwise likely theoretical dead end. I have argued that the standard focus on DRF as explanations for offending falls short of causal explanation and that these constructs suffer from a number of conceptual problems. By shifting the focus from lists of correlates to the mechanisms underpinning goal-directed behaviour, researchers can develop explanations which include normative aspects of human functioning, such as values, motivation, and reflection. They may then pinpoint the source of any impairments in functioning, but also view the behaviour in light of norms as well as contextual opportunities

and constraints. As well as a promising way forward for theory development, this offers up exciting new opportunities for treatment. I have suggested that agency theories may do this through adding depth to rehabilitation models such as the GLM. The PAM and future local models can do this through specifying the internal and external resources which are required for both the attainment of needs and reduction of risk, and thus which should be included as approach goals within an individual's GLP. It is hoped that this will eventually filter through to the design and delivery of treatment programmes which are more individualised and capable of motivating individuals towards prosocial change. After all, what good is a theory of offending if it does not have a positive impact on the lives of those who engage in and are affected by crime? We must not forget that, like us, persons who have committed crimes are motivated towards valued outcomes and experiences. Only when we understand them as humans first can we work alongside them to develop the resources required to live a good life.

References

Adshead, G., & Sarkar, S. P. (2009). Justice and welfare: Two ethical paradigms in forensic psychiatry. *Australian and New Zealand Journal of Psychiatry, 39,* 1011-1017. doi.org/10.1080/j.1440-1614.2005.01719.x

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory–Revised.* Toronto, ON: MultiHealth Systems.

Andrews, D. A., & Bonta, J. (2010). The psychology of criminal conduct (5th ed.). New Providence, NJ: LexisNexis.

Andrews, D. A., Bonta, J., & Wormith, J. (2004). *The Level of Service/Case Management Inventory (LS/CMI): User's manual.* Toronto, ON: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2010). The Level of Service (LS) assessment of adults and older adolescents. In R. K. Otto & K. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 199-225). New York, NY: Routledge.

Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of Variability in Estimates of Predictive Validity. *Criminal   Justice and Behavior, 38*(5), 413–432. doi:10.1177/0093854811401990

Aust, S. (2010). Is the good lives model of offender treatment relevant to sex offenders with a learning disability? *Journal of Learning Disabilities and Offending Behaviour*, *1*(3), 33-39. doi.org/10.5042/jldob.2010.0627

Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system.* Madison, WI: National Council on Crime & Delinquency.

Baird, C., Healy, T., Johnson, K., Bogie, A., Dankert, E. W., & Scharenbroch, C. (2013). *A comparison of risk assessment instruments in juvenile justice.* Madison, WI: National Council on Crime and Delinquency.

Baker, L. R. (2015). Human persons as social entities. *Journal of Social Ontology, 1,* 77-87. doi:10.1515/jso-2014-0037

Baker, L. R. (2016). Making sense of ourselves: Self-narratives and personal identity. *Phenomenology and the Cognitive Sciences, 1-15.* doi:10.1007/s11097-014-9358-y

Bar, M. (2011) (Ed.). *Predictions in the brain: Using our past to generate a future.* New York, NY: Oxford University Press.

Barnao, M., Ward, T., & Casey, S. (2015). Looking beyond the illness: Forensic service users' perceptions of rehabilitation. *Journal of Interpersonal Violence, 30*, 1025-1045. DOI: 10.1177/0886260514539764

Barnao, M., Ward, T., & Casey, S. (2016). Taking the good life to the institution: Forensic service users' perceptions of the Good Lives Model. *International Journal of Offender Therapy and Comparative Criminology* 60, 766-786. DOI: 10.1177/0306624X15570027

Barnett, G. D., Mandeville-Norden, R., & Rakestrow, J. (2014). The Good Lives Model or relapse prevention: What works better in facilitating change? *Sexual Abuse*, *26*(1), 3-33. DOI: 10.1177/1079063212474473

Barnett, G. D., Wakeling, H., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime & Law, 19*(1), 85–110. doi:10.1080/1068316x.2011.607820

Barnett, G., & Wood, J. L. (2008). Agency, relatedness, inner peace, and problem solving in sexual offending: How sexual offenders prioritize and operationalize their good lives conceptions. *Sexual Abuse*, *20*(4), 444-465. doi:10.1177/1079063208325202

Barrett, L. F., & Bar, M. (2011). See it with feeling: Affective predictions during object perceptions. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 107-121). New York, NY: Oxford University Press.

Bechtel, W. (2008). Mechanisms in cognitive psychology: What are the operations? *Philosophy of Science, 75*(5), 983-994. doi: 10.1086/594540

Beck, A. T. (1996). Beyond belief: A theory of modes, personality, and psychopathology. In P. Salkovskis (Ed.), *Frontiers of cognitive therapy* (pp. 1–25). New York, NY: Guilford.

Beech, A. R., & Craig, L. A. (2012). The current status of static and dynamic factors in sexual offender risk assessment. *Journal of Aggression, Conflict and Peace Research, 4*(4), 169-18. Doi.org.helicon.vuw.ac.nz/10.1108/17596591211270671.

Beech, A. R., & Mitchell, I. J. (2016). Intimacy Deficits/Attachment Problems in Sexual Offenders: Towards a Neurobiological Explanation. In A. R. Beech & T. Ward (Eds.). *The Wiley Handbook on the Theories, Assessment, & Treatment of Sexual Offending. Volume 1: Theories.* John Wiley & Sons.

Beech, A. R., Wakeling, H. C., Szumski, F., & Freemantle, N. (2016). Problems in the measurement of dynamic risk factors in sexual offenders. *Psychology, Crime & Law, 22*(1-2), 68–83. doi:10.1080/1068316x.2015.1109095

Beech, A. R., & Ward, T. (2004). The integration of etiology and risk in sexual offenders: A theoretical framework. *Aggression and violent behavior*, *10*(1), 31-63. doi.org/10.1016/j.avb.2003.08.002

Beggs, S. (2010). Within-treatment outcome among sexual offenders: A review. *Aggression and Violent Behavior, 15*(5), 369-379. doi:10.1016/j.avb.2010.06.005

Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology, 79(2), 182–192.* doi:10.1037/a0022900

Belsky, J., Schlomer, G. L., & Ellis, B. (2012). Beyond cumulative risk: Distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Developmental Psychology*, 48, 662–673. doi.org/10.1037/a004454

Bradford-Hill, A. (1965). The environment of disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58,* 295-300.

Bonta, J., & Andrews, D.A., (2017). *The psychology of criminal conduct* (6th ed.). New York, NY: Routledge.

Bowen, E., Gilchrist, E., & Beech, A. R. (2008). Change in treatment has no relationship with subsequent re-offending in UK domestic violence sample: A preliminary study. *International Journal of Offender Therapy and Comparative Criminology, 52*(5), 598-614. doi:10.1177/0306624X08319419

Brooks Holliday, S., Heilbrun, K., & Fretz, R. (2012). Examining improvements in criminogenic needs: The risk reduction potential of a structured re-entry program. *Behavioral Sciences & the Law, 30*(4), 431-447.DOI: 10.1002/bsl.2016

Carter, A. J., & Mann, R. E. (2016). Organizing principles for an integrated model of change for the treatment of sexual offending. In Beech, A. R., & Ward, T. (Vol. Eds.). *The Wiley handbook on the theories, assessment and treatment of sexual offending* (Vol. 1), (pp. 359-381). Chichester: Wiley. DOI:10.1002/9781118574003.wattso017

Casey, S. (2016). Dynamic risk and sexual offending: The conundrum of assessment. *Psychology, Crime & Law, 22*(1-2), 104-123. doi:10.1080/1068316X.2015.1111366

Christensen, W. (2012). Natural sources of normativity. *Studies in the History and Philosophy of Biological and Biomedical Sciences, 43,* 104-112. doi:10.1016/j.shpsc.2011.05.009

Chu, C. M., Koh, L. L., Zeng, G., & Teoh, J. (2015). Youth who sexual offended: Primary human goods and offense pathways. *Sexual Abuse: A Journal of Research and Treatment, 27*, 151-172. doi: 10.1177/1079063213499188

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind.* New York, NY: Oxford University Press.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 122, 155–159.

Cohen, T. H., Lowenkamp, C. T., & VanBenschoten, S. W. (2016). Does Change in Risk Matter? *Criminology & Public Policy, 15*(2), 263–296. doi:10.1111/1745-9133.12190

Collaborative Outcome Data Committee. (2007). *Sex offender treatment outcome research: Guidelines for evaluation (CODC Guidelines), Part 1: Introduction and overview* (Corrections User Report No 2007-002). Ottawa, Ontario: Public Safety Canada.

Cording, J. R., & Beggs Christofferson, S. M. (2017). Theoretical and practical issues for the measurement of protective factors. *Aggression and Violent Behavior, 32*, 45-54. doi:10.1016/j.avb.2016.12.007

Cording, J. R., Beggs Christofferson, S. M., & Grace, R. C. (2016). Challenges for the theory and application of dynamic risk factors. *Psychology, Crime & Law, 22*(1-2), 84-103. doi:10.1080/1068316x.2015.1111367

Coupland, R. B. A., & Olver, M. E. (2018). Assessing Dynamic Violence Risk in a High-Risk Treated Sample of Violent Offenders. *Assessment*. doi:10.1177/1073191118797440

Cuthbert, B. N., & Kozak, M. J. (2013). Constructing constructs for psychopathology: The NIMH Research Domain Criteria. *Journal of Abnormal Psychology, 122,* 3, 928-937. doi: 10.1037/a0034028

Damasio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. New York, NY: Pantheon.

Day, A. & Casey, S. (2009). Values in forensic and correctional psychology. *Aggression and Violent Behavior, 14,* 232-238. doi:10.1016/j.avb.2009.03.008

Day, A. & Ward, T. (2010). Offender rehabilitation as a value-laden process. *International Journal of Offender Therapy and Comparative Criminology, 54,* 289-306. DOI: 10.1177/0306624X09338284

Dealey, J. (2018). Moving beyond the risk paradigm: Using the good lives model with offenders in denial of sexual offending. European Journal of Probation, 10(1), 28-43. DOI: 10.1177/2066220318755530

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2009). *SAPROF. Guidelines for the assessment of protective factors for violence risk. **English** version.* Utrecht: Forum Educatief.

de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. *International Journal of Forensic Mental Health*, *10*(3), 171-177. doi.org/10.1080/14999013.2011.600230

de Vries Robbé, M. (2014). *Protective Factors: Validation of the Structured Assessment of Protective Factors for Violence Risk in Forensic Psychiatry* (doctoral thesis).

de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior, 39*(1), 53. doi:10.1037/lhb0000089

de Vries Robbé, M., Mann, R. E., Maruna, S., & Thornton, D. (2015). An exploration of protective factors supporting desistance from sexual offending. *Sexual Abuse, 27*(1), 16-33. doi:10.1177/1079063214547582

Di Lorito, C., Völlm, B., & Dening, T. (2018). Psychiatric disorders among older prisoners: a systematic review and comparison study against older people in the community. *Aging & mental health*, *22*(1), 1-10. doi.org/10.1080/13607863.2017.1286453

Douglas, H. E. (2009). *Science, policy, and the value-free ideal.* Pittsburgh, PA: University of Pittsburgh Press.

Douglas, H. E. (2014). The moral terrain of science. *Erkenntnis, 79,*961-979. DOI: 10.1007/s10670-013-9538-0

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. Psychology, Public Policy, and Law, 11(3), 347–383. doi:10.1037/1076–8971.11.3.347

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, *42*, 449.

Durrant, R. (2017). Why do protective factors protect? An evolutionary developmental perspective. *Aggression and Violent Behavior, 32*, 4-10. doi:10.1016/j.avb.2016.12.002

Durrant, R., & Ward, T. (2015). *Evolutionary criminology: Towards a comprehensive explanation of crime and its management.* San Diego, CA: Academic Press.

Duwe, G., & Rocque, M. (2016). A Jack of all trades but a master of none? Evaluating the performance of the Level of Service Inventory–Revised (LSI-R) in the assessment of risk and need. *Corrections, 1*(2), 81-106. doi.org/10.1080/23774657.2015.1111743

Farrall, S. (2005). On the existential aspects of desistance from crime. *Symbolic Interaction, 28*(3), 367-386. doi: 10.1525/si.2005.28.issue-3

Farrington, D. P. (2016). Risk, promotive, and protective factors in youth offending: Results from the Cambridge study in delinquent development. *Journal of Criminal Justice, 45*, 63-70. doi.org/10.1016/j.jcrimjus.2016.02.014

Farrington, D. P., Loeber, R., & Ttofi, M. M. (2012). Risk and protective factors for offending. In B. C. Welsh & D. P. Farrington (Eds.), *The Oxford handbook of crime prevention*, (pp. 46-69). New York: Oxford university Press.

Finkelhor. D. (1984). *Child sexual abuse: New theory and research.* New York: The Free Press.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press (Taylor & Francis).

Fletcher, G., Simpson, J. A., Campbell, L., & Overall, N. C. (2013). *The science of intimate relationships*. Sussex, UK: Wiley-Blackwell.

Fortune, C. A. (2017). The Good Lives Model: A strength-based approach for youth offenders. *Aggression and Violent Behavior, 38,* 21-30. doi.org/10.1016/j.avb.2017.11.003

Fortune, C-A., & Heffernan, R. (2019). The Psychology of Criminal Conduct: Criticisms and future directions. *Psychology, Crime & Law*, 25(6), 659-674. doi.org/10.1080/1068316X.2018.1560445

Fortune, C.-A., & Ward, T. (2017). Problems in protective factor research and practice. *Aggression and Violent Behavior, 32*, 1-3. doi:10.1016/j.avb.2016.12.008

Gannon, T. A. (2016). Forensic psychologists should use the behavioral experiment to facilitate cognitive change in clients who have offended. *Aggression and violent behavior*, 27, 130-141. doi.org/10.1016/j.avb.2016.02.008

Gannon, T., King, T., Miles, H., Lockerbie, L., & Willis, G. M. (2011). Good lives sexual offender treatment for mentally disordered offenders. *British Journal of Forensic Practice*, 13(3), 153-168. DOI 10.1108/14636641111157805

Gannon, T. A., Olver, M. E., Mallion, J. S., & James, M. (2019). Does specialized psychological treatment for offending reduce recidivism? A meta-analysis examining staff and program variables as predictors of treatment effectiveness. *Clinical Psychology Review, 73,* 101752. https://doi.org/10.1016/j.cpr.2019.101752

Gannon, T., & Ward, T. (2014). Where has all the Psychology Gone? A Critical Review of Evidence-Based Psychological Practice in Correctional Settings. *Aggression and Violent Behavior, 19,* 435-436. doi.org/10.1016/j.avb.2014.06.006

Giguère, G., & Lussier, P. (2016). Debunking the psychometric properties of the LS\CMI: An application of item response theory with a risk assessment instrument. *Journal of Criminal Justice, 46,* 207–218. doi:10.1016/j.jcrimjus.2016.05.005

Glaser, B. (2003). Therapeutic jurisprudence: An ethical paradigm for therapists in sex offender treatment programs. *Western Criminology Review* 4(2), 143-154.

Grady, M. D., Brodersen, M., & Abramson, J. M. (2011). The state of psychological measures for adult sexual offenders. *Aggression and Violent Behavior, 16*(3), 227-240. doi:10.1016/j.avb.2011.03.001

Grady, M. D., Edwards Jr, D., & Pettus-Davis, C. (2017). A longitudinal outcome evaluation of a prison-based sex offender treatment program. *Sexual Abuse*, *29*(3), 239-266. DOI: 10.1177/1079063215585731

Greiner, L. E., Law, M. A., & Brown, S. L. (2015). Using dynamic factors to predict recidivism among women: A four-wave prospective study. *Criminal Justice and Behavior, 42*(5), 457-480. DOI: 10.1177/0093854814553222

Haig, B. (2014). *Investigating the Psychological World*. Cambridge, Massachussets: The MIT Press.

Hall, G. C. N., & Hirschman, R. (1992). Sexual aggression against children: A conceptual perspective of etiology. *Criminal Justice and Behavior*, *19*(1), 8-23. doi/pdf/10.1177/0093854892019001003

Hannah-Moffat, K. (2009). Gridlock or mutability: Reconsidering "gender" and risk assessment. *Criminology & Public Policy, 8*(1), 209–219. doi:10.1111/j.1745-9133.2009.00549.x

Hannah-Moffat, K. (2016). A conceptual kaleidoscope: Contemplating 'dynamic structural risk' and an uncoupling of risk from need. *Psychology, Crime & Law, 22*(1-2), 33-46. doi.org/10.1080/1068316X.2015.1114115

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie Canadienne, 50*(3), 172–182. doi:10.1037/a0015726

Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, *36*(9), 865-891. DOI: 10.1177/0093854809338545

Hanson, R. K., Bourgon, G., McGrath, R. K., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017). A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language. New York, NY: The Council of State Governments Justice Center.

Hanson, K., R., & Harris, A., J.R. (2000). Where should we intervene? Dynamic Predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*(1), 6-35. doi/pdf/10.1177/0093854800027001002

Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse, 13*(2), 105-122. doi.org/10.1177/107906320101300204

Hanson, R. K., Harris, A. J., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project (Corrections research user report 2007-05)*. Ottawa, Ontario: Public Safety Canada. Retrieved from www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/index-en.aspx

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1-21. doi.org/10.1037/a0014421

Harkins, L., Flak, V. E., Beech, A. R., & Woodhams, J. (2012). Evaluation of a community-based sex offender treatment program using a good lives model approach. *Sexual Abuse*, *24*(6), 519-543. DOI: 10.1177/1079063211429469

Harris, D. A., Pedneault, A., & Willis, G. (2017). The pursuit of primary human goods in men desisting from sexual offending. *Sexual Abuse*, 00(0), 1-23. DOI: 1079063217729155

Hart, S. D. (2009). Evidence-based assessment of risk for sexual violence. *Chapman Journal of Criminal Justice, 1*, 143-165.

Hart, S. D., & Logan, C. (2011). Formulation of violent risk using evidence-based assessments: The Structured Professional Judgment Approach. In P. Sturmey & M.

McMurran (Eds.), *Forensic case formulation* (pp. 3-32). West Sussex, UK: John
Wiley & Sons.

Hart, S. D., Sturmey, P., Logan, C., & McMurran, M. (2011). Forensic case formulation.
International Journal of Forensic Mental Health, 10, 118-126. DOI:
10.1080/14999013.2011.577137

Heffernan, R., & Ward, T. (2015). The conceptualization of dynamic risk factors in child sex
offenders: An agency model. *Aggression and Violent Behavior,24,* 250-260.
doi.org/10.1016/j.avb.2015.07.001

Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective
factors. *Aggression and Violent Behavior, 37,* 129-141.
doi.org/10.1016/j.avb.2017.10.003

Heffernan, R., & Ward, T. (2019). Dynamic risk factors, protective factors, and value-laden
practices. *Psychiatry, Psychology and Law, 26*(2), 312-328.
doi.org/10.1080/13218719.2018.1506721

Heffernan, R., Ward, T., Vandevelde, S., & Van Damme, L. (2019). Dynamic risk factors and
constructing explanations of offending: The Risk-Causality Method. *Aggression and
Violent Behavior, 44*, 47-56. doi:10.1016/j.avb.2018.11.009

Heffernan, R., Wegerhoff, D., & Ward, T. (2019). Dynamic risk factors: Conceptualization,
measurement, and evidence. *Aggression and Violent Behavior (48),* 6-16.
doi.org/10.1016/j.avb.2019.06.004

Helmus, L. M. (2018). Sex offender risk assessment: Where are we and where are we going?
Current psychiatry reports, 20(6), 46. doi:10.1007/s11920-018-0909-8

Hendry, M. C. (2013). *Hostile attribution bias as a dynamic risk factor in civil psychiatric
patients and criminal offenders: Change over time and relationship to violence and
recidivism* (Doctor of Philosophy). Simon Fraser University, Canada.

Hildebrand, M., & De Ruiter, C. (2012). Psychopathic traits and change on indicators of
dynamic risk factors during inpatient forensic psychiatric treatment. *International
Journal of Law and Psychiatry, 35*(4), 276-288. doi:10.1016/j.ijlp.2012.04.001

Hohwy, J. (2013). *The predictive mind.* New York, NY: Oxford University Press.

Hough, M. (2010). Gold standard or fool's gold? The pursuit of certainty in experimental
criminology. *Criminology & Criminal Justice*, *10*(1), 11-22. DOI:
10.1177/1748895809352597

Howard, M., & Van Doorn, G., (2018). Within-treatment change in antisocial attitudes and reoffending in a large sample of custodial and community offenders. *Law and Human Behavior, 42*(4), 321-335. doi:10.1037/lhb0000290

Ibrahim, S., & Alkire, S. (2007). Agency and empowerment: A proposal for internationally comparable indicators. *Oxford Development Studies, 35*:4, 379-403. doi: 10.1080/13600810701701897

Johnson, M. (2014). *Morality for humans: Ethical understanding from the perspective of cognitive science.* Chicago, IL: University of Chicago Press.

Jolliffe, D., & Farrington, D. P. (2007, December). *A systematic review of the national and international evidence on the effectiveness of interventions with violent offenders.* Series 16/07. London: Ministry of Justice Research.

Jung, S., & Gulayets, M. (2011). Using clinical variables to evaluate treatment effectiveness in programmes for sexual offenders. *Journal of Sexual Aggression, 17*(2), 166-180. DOI: 10.1080/13552601003802238

Kingston, D. A., & Olver, M. E. (2018). Psychometric Examination of Treatment   Change Among Mentally Disordered Offenders: A Risk–Needs Analysis.   *Criminal Justice and Behavior, 45*(2), 153–172. doi:10.1177/0093854817743539

Klepfisz, G. (2018). Violent offenders' within-treatment change in dynamic risk and protective factors: associations with violent recidivism (Doctoral dissertation, Swinburne University of Technology). Retrieved from https://researchbank.swinburne.edu.au/items/12ac4bf3-4e82-485a-8e79-68cdf0fe42bc/1/ (2018)

Klepfisz, G., Daffern, M., & Day, A. (2016). Understanding dynamic risk factors for violence. *Psychology, Crime & Law, 22*(1-2), 124–137. doi:10.1080/1068316X.2015.1109091

Klepfisz, G., O'Brien, K., & Daffern, M. (2014). Violent offenders' within-treatment change in anger, criminal attitudes, and violence risk: Associations with violent recidivism. *International Journal of Forensic Mental Health, 13*(4), 348-362. doi: 10.1080/14999013.2014.951107

Kozar, C. J., & Day, A. (2012). The therapeutic alliance in offending behavior programs: A necessary and sufficient condition for change?. *Aggression and Violent Behavior, 17*(5), 482-487. doi.org/10.1016/j.avb.2012.07.004

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D.   J. (1997). Coming to terms with the terms of risk. *Archives of General   Psychiatry*, *54,* 337-343. doi:10.1001/archpsyc.1997.01830160065009

Kroner, D. G., Polaschek, D. L., Serin, R. C., & Skeem, J. L. (2017). An exploration of the symmetry between crime-causing and crime-reducing factors: Implications for delivery of offender services. *Psychological Services, 16*(2), 329-339. DOI:10.1037/ser0000199

Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behavior: the structure of criminal risk. *Int J Law Psychiatry, 28*(4), 360-374. doi:10.1016/j.ijlp.2004.01.011

Kroner, D. G., & Yessine, A. K. (2013). Changing risk factors that impact recidivism: In search of mechanisms of change. *Law and Human Behavior, 37*(5), 321–336. doi: 10.1037/lhb0000022

Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The Importance of Reassessment: How Changes in the LSI-R Risk Score Can Improve the Prediction of Recidivism. *Journal of Offender Rehabilitation, 53*(2), 116–128. doi:10.1080/10509674.2013.868389

Langlands, R. L., Ward, T., & Gilchrist, E. (2009). Applying the good lives model to male perpetrators of domestic violence. *Behaviour Change*, *26*(2), 113-129. doi.org/10.1375/bech.26.2.113

Latessa, E. J., & Lovins, B. (2010). The Role of Offender Risk Assessment: A Policy Maker Guide. *Victims & Offenders, 5*(3), 203–219. doi:10.1080/15564886.2010.485900

Laws, D. R., & Ward, T. (2011). *Desistance from sex offending: Alternatives to throwing away the keys*. New York. NY: Guilford Press.

Leaming, N., & Willis, G. M., (2016). The good lives model: New avenues for Māori rehabilitation? *Sexual Abuse in Australia and New Zealand, 7*, 59-69. ISSN: 1833-8488

Lilienfeld, S. O. (2014). The Research Domain Criteria (RDoC): An analysis of methodological and conceptual challenges. *Behaviour Research and Therapy, 62*, 129-139. doi.org/10.1016/j.brat.2014.07.019

Lindsay, W. R., Ward, T., Morgan, T., & Wilson, I. (2007). Self-regulation of sex offending, future pathways and the good lives model: Applications and problems. *Journal of Sexual Aggression*, *13*(1), 37-50. doi.org/10.1080/13552600701365613

Loney, D. M., & Harkins, L. (2018). Examining the good lives model and antisocial behaviour. *Psychology, Crime & Law, 24*(1), 38-51. doi.org/10.1080/1068316X.2017.1371304

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine*, *43*(2), p8-23. doi.org/10.1016/j.amepre.2012.04.029

Lovins, B. K., Latessa, E. J., May, T., & Lux, J. (2018). Validating the Ohio Risk Assessment System community supervision tool with a diverse sample from Texas. *Corrections*, *3*(3), 186-202. doi.org/10.1080/23774657.2017.1361798

Magnus, P. D. (2012). *Scientific inquiry and natural kinds: From planets to mallards.* Basingstoke, UK: Palgrave MacMillan.

Maiese, M. (2011). *Embodiment, emotion, and cognition.* Basingstoke, UK: Palgrave Macmillan.

Maiese, M. (2016). *Embodied selves and divided minds.* New York, NY: Oxford University Press.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: some proposals on the nature of psychologically meaningful risk factors. *Sex Abuse, 22*(2), 191-217. doi:10.1177/1079063210366039

Mann, R. E., Webster, S. D., Schofield, C., & Marshall, W. L. (2004). Approach versus avoidance goals in relapse prevention with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *16*(1), 65-75. doi/pdf/10.1177/107906320401600105

Marshall, W. L., & Barbaree, H. E. (1990). *An integrated theory of sexual offending*. New York: Springer.

Marshall, W. L., & Marshall, L. E. (2007). The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A Journal of Research and Treatment, 19,* 175-191-260. doi/pdf/10.1177/107906320701900207

Marshall, W. L., & Marshall, L. E. (2017). The treatment of adult male sexual offenders. In D. P. Boer (Ed.). *Theories, assessment, and treatment of sexual offending* (pp. 1227-1243). Chichester, England: Wiley Blackwell.

Marshall, W. L., Marshall, L. E., Serran, G. A., & O'Brien, M. D. (2011). *Rehabilitating sexual offenders: A strength-based approach*. Washington, DC: American Psychological Association.

Marshall, W. L., & Serran, G. A. (2004). The role of the therapist in offender treatment. *Psychology, Crime & Law, 10*(3), 309-320. doi.org/10.1080/10683160410001662799

Martinson, R. (1974). What works? Questions and answers about prison reform. *The public interest*, *35*, 22. ISSN: 0033-3557

Maruna, S. (2011). Reentry as a rite of passage. *Punishment & Society*, 13(1), 3-28. DOI: 10.1177/1462474510385641

Mastromanno, B., Brookstein, D. M., Ogloff, J. R. P., Campbell, R., Chu, C. M., & Daffern, M. (2017). Assessing change in dynamic risk factors in forensic   psychiatric inpatients: relationship with psychopathy and recidivism. *The   Journal of Forensic Psychiatry & Psychology, 29*(2), 323–336. doi:10.1080/14789949.2017.1377277

McAlinden, A. M., Farmer, M., & Maruna, S. (2017). Desistance from sexual offending: Do the mainstream theories apply? *Criminology & Criminal Justice*, *17*(3), 266-283. doi.org/10.1177/1748895816670201

McMurran, M., & Theodosi, E. (2007). Is treatment non-completion associated with increased reconviction over no treatment? *Psychology, Crime & Law*, *13*(4), 333-343. doi.org/10.1080/10683160601060374

McNeill, F. (2012). Four forms of 'offender'rehabilitation: Towards an interdisciplinary perspective. *Legal and Criminological Psychology, 17*(1), 18-36. DOI:10.1111/j.2044-8333.2011.02039.x

Mitchell, S., D. (2003). *Biological Complexity and Integrative Pluralism.* Cambridge, UK: Cambridge University Press.

Morgan, R. D., Kroner, D. G., Mills, J. F., Serna, C., & McDonald, B. (2013). Dynamic risk assessment: A validation study. *Journal of Criminal Justice, 41*(2), 115- 124. doi:10.1016/j.jcrimjus.2012.11.004

Murphy, D. (2016). Validity, realism, and normativity. In P. Zachar, D. S. Stoyanov, M. Aragona, & A. Jablensky. (Eds.), *Alternative perspectives on psychiatric validation* (p.60-75). Oxford, UK: Oxford University Press.

Nee, C., & Meenaghan, A. (2006). Expert decision-making in burglars. *British Journal of Criminology*, 46(5), 935–949, doi.org/10.1093/bjc/az1013

Neisser, J. (2015). *The science of subjectivity.* Basingstoke, UK: Palgrave MacMillan.

Newsome, J., & Cullen, F. T. (2017). The risk-need-responsivity model revisited: Using biosocial criminology to enhance offender rehabilitation. *Criminal Justice and Behavior, 44*(8), 1030–1049. doi:10.1177/0093854817715289

Nunes, K. L., Pedneault, C. I., Filleter, W. E., Maimone, S., Blank, C., & Atlas, M. (2019). "I know correlation doesn't prove causation, but...": Are we jumping to unfounded conclusions about the causes of sexual offending?. *Sexual Abuse, 31*(2), 220-236. doi.org/10.1177/1079063217729156

O'Brien, K., & Daffern, M. (2017). Treatment gain in violent offenders: The relationship between proximal outcomes, risk reduction and violent recidivism. *Psychiatry, Psychology and Law, 24*(2), 244–258. doi:10.1080/13218719.2016.1209804

Olver, M. E., Beggs Christofferson, S. M., & Wong, S. C. (2015). Evaluation and applications of the clinically significant change method with the Violence Risk Scale-Sexual Offender version: Implications for risk-change communication. *Behavioral Sciences and the Law, 33*(1), 92-110. doi:10.1002/bsl.2159

Olver, M. E., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C. (2014). A psychometric examination of treatment change in a multisite sample of treated Canadian federal sexual offenders. *Law and Human Behavior, 38*(6), 544-559. doi:10.1037/lhb0000086

Olver, M. E., Lewis, K., & Wong, S. C. (2013). Risk reduction treatment of high-risk psychopathic offenders: The relationship of psychopathy and treatment change to violent recidivism. *Personality Disorders: Theory, Research, and   Treatment, 4*(2), 160. doi:10.1037/a0029769

Olver, M. E., Marshall, L. E., Marshall, W. L., & Nicholaichuk, T. P. (2018). A long-term outcome assessment of the effects on subsequent reoffense rates of a prison-based CBT/RNR sex offender treatment program with strength-based elements. *Sexual Abuse, 32*(2),127-153. Doi:10.1177/1079063218807486

Olver, M. E., Mundt, J. C., Thornton, D., Beggs Christofferson, S. M., Kingston, D. A., Sowden, J. N., ... & Wong, S. C. (2018). Using the Violence Risk Scale-Sexual Offense version in sexual violence risk assessments: Updated risk categories and recidivism estimates from a multisite sample of treated sexual offenders. *Psychological Assessment, 30*(7), 941. doi.org/10.1037/pas0000538

Olver, M. E., Nicholaichuk, T. P., Kingston, D. A., & Wong, S. C. P. (2014). A multisite examination of sexual violence risk and therapeutic change. *Journal of Consulting and Clinical Psychology, 82*(2), 312–324. doi:10.1037/a0035340

Olver, M. E., Nicholaichuk, T. P., & Wong, S. C. (2014). The predictive and convergent validity of a psychometric battery used to assess sexual offenders   in a treatment

programme: An 18-year follow-up. *Journal of Sexual   Aggression, 20*(2), 216-239. doi:10.1080/13552600.2013.816791

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, *26*(1), 156. doi.org/10.1037/a0035080

Olver, M. E., & Wong, S. C. P. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior, 38*(2), 113–126. doi:10.1177/0093854810389534

Olver, M. E., & Wong, S. C. P. (2013). Treatment programs for high risk sexual offenders: Program and offender characteristics, attrition, treatment change and recidivism. *Aggression and Violent Behavior, 18*(5), 579-591. doi:10.1016/j.avb.2013.06.002

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*(3), 318-329. doi:10.1037/1040-3590.19.3.318

Polaschek, D. L. L. (2012). An appraisal of the risk–need–responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology, 17*(1), 1-17. DOI:10.1111/j.2044-8333.2011.02038.x

Polaschek, D. L. (2016). Desistance and dynamic risk factors belong together. *Psychology, Crime & Law*, *22*(1-2), 171-189. doi.org/10.1080/1068316X.2015.1114114

Polaschek, D. L L. (2017). Protective factors, correctional treatment and desistance. *Aggression and Violent Behavior, 32,* 64-70. doi.org/10.1016/j.avb.2016.12.005

Purvis, M. (2010). *Seeking a Good Life: Human Goods and Sexual Offending*. Published PhD Manuscript. Lambert Academic Press, Germany.

Purvis, M., Ward, T., & Willis, G. (2011). The good lives model in practice: Offence pathways and case management. *European Journal of Probation, 3*(2), 4-28. doi/pdf/10.1177/206622031100300202

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, *57*(3), 316. doi.org/10.1111/j.1939-0025.1987.tb03541.x

Rutter, M. (2007). Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science*, *2*(4), 377-395. doi/pdf/10.1111/j.1745-6916.2007.00050.x

Sadler, J. Z. (2005). *Values and psychiatric diagnosis.* New York, NY: Oxford University Press.

Sampson, R. J., & Laub, J. H. (2005). A life-course view of the development of crime. *The ANNALS of the American Academy of Political and Social Science, 602*(1), 12-45. doi:10.1177/0002716205280075

Schlager, M. D., & Pacheco, D. (2011). An Examination of Changes in LSI-R Scores Over Time: Making the Case for Needs-Based Case Management. *Criminal Justice and Behavior, 38*(6), 541–553. doi:10.1177/0093854811402300

Schmidt, S., Heffernan, R., & Ward, T. (2020). Why we cannot explain cross-cultural differences in risk assessment. *Aggression and Violent Behavior, 50,* 1-12. DOI:10.1016/j.avb.2019.101346

Schmucker, M., & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: An international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology, 11*(4), 597-630. doi:10.1007/s11292-015-9241-z

Seligman, M. E., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. Perspectives on psychological science, 8(2), 119-141. DOI: 10.1177/1745691612474317

Seligman, M. E. P., Railton, P., Baumeister, R. R., & Sripada, C. (2016). *Homo Prospectus.* New York, NY: Oxford University Press.

Serin, R. C., Chadwick, N., & Lloyd, C. D. (2016). Dynamic risk and protective factors. *Psychology, Crime & Law*, *22*(1-2), 151-170. doi.org/10.1080/1068316X.2015.1112013

Serin, R. C., & Lloyd, C. D. (2009). Examining the process of offender change: The transition to crime desistance. Psychology, Crime & Law, 15(4), 347-364. doi.org/10.1080/10683160802261078

Serin, R. C., & Hanby, L. J. (2016). Client-Based Assessment of Need and Change. *The Wiley Handbook on the Theories, Assessment and Treatment of Sexual Offending*, 1575–1592. doi:10.1002/9781118574003.wattso077

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the holy grail in assessing offender change. *Aggression and Violent Behavior, 18*(1), 32-53. doi:10.1016/j.avb.2012.09.002

Shepherd, S. M., & Lewis-Fernandez, R. (2016). Forensic risk assessment and cultural diversity: Contemporary challenges and future directions. *Psychology, Public Policy, and Law*, *22*(4), 427. doi.org/10.1037/law0000102

Simons, D. A., McCullar, B., & Tyler, C. (2008, October). *The Utility of the Self- Regulation Model to re-integration planning*. Paper presented at the 27th Annual Association for the Treatment of Sexual Abusers Research and Treatment Conference.

Smallbone, S., & Cale, J. (2016). Situational Theories. In A. R. Beech & T. Ward (Eds.). *The Wiley Handbook on the Theories, Assessment, & Treatment of Sexual Offending. Volume 1: Theories.* John Wiley & Sons.

Smallbone, S. W., & Dadds, M. R. (2000). Attachment and coercive sexual behavior. *Sexual Abuse: A Journal of Research and Treatment, 12,* 3-15. doi/pdf/10.1177/107906320001200102

Strauss-Hughes, A., Heffernan, R., & Ward, T. (2019). A cultural–ecological perspective on agency and offending behavior. *Psychiatry, Psychology and Law*, DOI: 10.1080/13218719.2019.1644250

Sterelny, K. (2012). *The evolved apprentice: How evolution made humans unique.* Cambridge, MA: MIT Press.

Stinson, J. D. Becker, J. V. & McVay, L. A. (2016). Multimodal Self-Regulation Theory of Sexual Offending. In A. R. Beech & T. Ward (Eds.). *The Wiley Handbook on the Theories, Assessment, & Treatment of Sexual Offending. Volume 1: Theories.* John Wiley & Sons.

Sturgess, D., Woodhams, J., & Tonkin, M. (2016). Treatment engagement from the perspective of the offender: reasons for noncompletion and completion of treatment—a systematic review. *International Journal of Offender Therapy and Comparative Criminology*, *60*(16), 1873-1896.

Sturmey, P., & McMurran, M. (Eds). (2011). *Forensic case formulation*. Oxford: Wiley-Blackwell.

Suddendorf, T. (2013). *The Gap: The science of what separates us from animals.* New York, NY: Basic Books.

Sullivan, J. A. (2019). Achieving Cumulative Progress in Understanding Crime: Some Insights from the Philosophy of Science. *Psychology, Crime & Law*, 1-29. doi:10.1080/1068316X.2019.1572755

Tan, L., & Grace, R. C. (2008). Social Desirability and Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment, 20*(1), 61–87. doi:10.1177/1079063208314820

Tappolet, C. & Rossi, M. (2016). What is value: Where does it come from? A philosophical perspective. In In T. Brosch & D. Sandler (Eds.), *Handbook of value: Perspectives from economics, neuroscience, philosophy, psychology, and sociology* (pp.3-22.). Oxford, UK: Oxford University Press.

Taxman, F. S. (2017). (Ed.). *Handbook on risk and need assessment: Theory and practice.* New York, NY: Routledge.

Taxman, F. S., & Caudy, M. S. (2015). Risk tells us who, but not what or how: Empirical assessment of the complexity of criminogenic needs to inform correctional programming. *Criminology & Public Policy, 14*(1), 71-103. doi/pdf/10.1111/1745-9133.12116

Taylor, E. (2017). 'I should have been a security consultant': The good lives model and residential burglars. *European Journal of Criminology, 14*(4), 434-450. doi.org/10.1177/1477370816661743

Thagard, P. (1999). *How scientists explain disease.* Princeton, NJ: Princeton University Press.

Thakker, J., & Ward, T. (2010). The good lives model and the treatment of substance abusers. *Behaviour Change*, *27*(3), 154-75. doi:10.1375/bech.27.3.154

Thomas, J. G., & Sharp, P. B. (2019). Mechanistic science: A new approach to comprehensive psychopathology research that relates psychological and biological phenomena. Clinical Psychological Science, 7(2), 196-215. doi.org/10.1177/2167702618810223

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of the mind.* MA, USA: Belknap-Harvard.

Thornton, D. (2016). Developing a theory of dynamic risk. *Psychology, Crime & Law, 22*(1-2), 138-150. doi:10.1080/1068316x.2015.1109092

Thornton, D., Kelley, S. M., & Nelligan, K. E. (2017). Protective factors and mental illness in men with a history of sexual offending. *Aggression and Violent Behavior, 32*, 29-36. doi:10.1016/j.avb.2016.12.003

Thornton, D., & Laws, D. R. (2009). *Cognitive approaches to the assessment of sexual interest in sexual offenders*. Hoboken, NJ: Wiley.

Thornton, D., & Knight, R. A. (2015). Construction and validation of SRA-FV need assessment. *Sexual Abuse*, *27*(4), 360-375. DOI: 10.1177/1079063213511120

Tierney, D. W., & McCabe, M. P. (2001). An evaluation of self-report measures of cognitive distortions and empathy among Australian sex offenders. *Archives of Sexual Behavior, 30*(5), 495-519. doi:10.1023/A:1010239217517

Van Damme, L., Hoeve, M., Vermeiren, R., Vanderplasschen, W., & Colins, O. F. (2016). Quality of life in relation to future mental health problems and offending: Testing the good lives model among detained girls. *Law and Human Behavior, 40*, 285-294. doi.org/10.1037/lhb0000177

van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment, 30*(2), 179. doi:10.1037/pas0000454

Vose, B., Lowenkamp, C. T., Smith, P., & Cullen, F. T. (2009). Gender and the Predictive Validity of the LSI-R: A Study of Parolees and Probationers. *Journal of Contemporary Criminal Justice, 25*(4), 459–471. doi:10.1177/1043986209344797

Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior, 40*(12), 1383-1396. doi:10.1177/0093854813508916

Wainwright L, & Nee, C. (2014). The good lives model – new directions for preventative practice with children. *Psychology, Crime & Law 20*(2), 166–182. doi.org/10.1080/1068316X.2013.770851

Wakeling, H., Beech, A. R., & Freemantle, N. (2013). Investigating treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK. *Psychology, Crime & Law, 19*(3), 233-252. doi:10.1080/1068316X.2011.626413

Wallace, J. D. (2009). *Norms and practices*. Ithaca, NY: Cornell University Press.

Walsh, D. M. (2015). *Organisms, agency, and evolution.* Cambridge, UK: Cambridge University Press.

Ward, T. (2002a). Good lives and the rehabilitation of sexual offenders: Promises and problems. *Aggression and Violent Behavior, 7*, 513-528. doi.org/10.1016/S1359-1789(01)00076-3

Ward, T. (2002b). The management of risk and the design of good lives. *Australian Psychologist, 37*, 172-179. doi.org/10.1080/00050060210001706846

Ward, T. (2013). Addressing the dual relationship problem in forensic and correctional practice. *Aggression and Violent Behavior, 18*(1). 92-100. doi.org/10.1016/j.avb.2012.10.006

Ward, T. (2014). The explanation of sexual offending: from single factor theories to integrative pluralism. *Journal of Sexual Aggression, 20*(2), 130-141. doi:10.1080/13552600.2013.870242

Ward, T. (2016). Dynamic risk factors: scientific kinds or predictive constructs. *Psychology, Crime & Law, 22*(1-2), 2-16. doi:10.1080/1068316x.2015.1109094

Ward, T. (2017). Prediction and agency: The role of protective factors in correctional rehabilitation and desistance. *Aggression and Violent Behavior, 32*, 19-28. doi:10.1016/j.avb.2016.11.012

Ward, T. (2019). Why theory matters in correctional psychology. *Aggression and Violent Behavior, 48,* 36-45. doi.org/10.1016/j.avb.2019.08.015

Ward, T., & Beech, A. R. (2006). An integrated theory of sexual offending. *Aggression and Violent Behavior, 11,* 44-63. doi.org/10.1016/j.avb.2005.05.002

Ward, T., & Beech, A. R. (2015). Dynamic risk factors: a theoretical dead-end? *Psychology, Crime & Law, 21*(2), 100-113. doi:10.1080/1068316x.2014.917854

Ward, T., & Birgden, A. (2007). Human rights and clinical correctional practice. *Aggression and Violent Behavior, 12*, 628-643. doi:10.1016/j.avb.2007.05.001

Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law, 10*(3), 243-257. doi.org/10.1080/10683160410001662744

Ward, T. & Carter, E. (2019). The classification of offending and crime related problems: A functional perspective. Psycho*logy, Crime, & Law, 25*(6), 542-560. doi:10.1080/1068316X.2018.1557182

Ward, T., & Fisher, D. D. (2005). New ideas in the treatment of sexual offenders. In W. L. Marshall, Y. Fernandez, L. Marshall and G. A. Serran (Eds.), *Sexual offender treatment: Issues and controversies* (pp. 143-148). Chichester, UK: John Wiley and Sons Ltd.

Ward, T., & Fortune, C.-A. (2016a). The role of dynamic risk factors in the explanation of offending. *Aggression and Violent Behavior, 29*, 79-88. doi:10.1016/j.avb.2016.06.007

Ward, T., & Fortune, C.-A. (2016b). From dynamic risk factors to causal processes: a methodological framework. *Psychology, Crime & Law, 22*(1-2), 190-202. doi:10.1080/1068316x.2015.1117080

Ward, T., & Heffernan, R. (2017). The role of values in forensic and correctional rehabilitation. *Aggression and Violent Behavior, 37,* 42-51. doi.org/10.1016/j.avb.2017.09.002

Ward, T., & Hudson, S. M. (1998a). The construction and development of theory in the sexual offending area: A metatheoretical framework. *Sexual Abuse: A Journal of Research and Treatment, 10*(1), 47-63. doi/pdf/10.1177/107906329801000106

Ward, T., & Hudson, S. M. (1998b). A model of the relapse process in sexual offenders. *Journal of Interpersonal Violence*, 13, 700-725. doi/pdf/10.1177/088626098013006003

Ward, T., Hudson, S., & Marshall, W. L. (1995). Cognitive distortions and affective deficits in sex offenders: a cognitive deconstructionist interpretation. *Sexual Abuse, 7*(1), 67-83. doi/pdf/10.1177/107906329500700107

Ward, T., Hudson, S., & Marshall, W. L. (1996). Attachment style in sex offenders: A preliminary study. *Journal of Sex Research, 33*, 17–26.

Ward, T., & Keenan, T. (1999). Child molesters' implicit theories. *Journal of Interpersonal Violence*, *14*(8), 821-838. doi/pdf/10.1177/088626099014008003

Ward, T., Mann, R. E., & Gannon, T. A. (2007). The good lives model of offender rehabilitation: Clinical implications. *Aggression and Violent Behavior*, *12*(1), 87-107. doi:10.1016/j.avb.2006.03.004

Ward, T., & Maruna, S. (2007). *Rehabilitation: Beyond the risk paradigm.* London, UK: Routledge.

Ward, T., Melser, J., & Yates, P. M. (2007). Reconstructing the Risk-Need-Responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior, 12*(2), 208-228. doi.org/10.1016/j.avb.2006.07.001

Ward, T., & Moreton, G. (2008). Moral Repair with Offenders: Ethical Issues Arising From Victimization Experiences. *Sexual Abuse: A Journal of Research and Treatment, 20,* 305-322. doi: 10.1177/1079063208322423

Ward, T., Polaschek, D. L., & Beech, A. R. (2006). *Theories of sexual offending* (Vol. 21). Chichester, UK: Wiley.

Ward, T., & Siegert, R. J. (2002). Toward a comprehensive theory of child sexual abuse: A theory knitting perspective. *Psychology, Crime & Law, 8*(4), 319-351. doi:10.1080/10683160208401823

Ward, T., & Stewart, C. A. (2003). The treatment of sex offenders: Risk management and good lives. *Professional Psychology: Research and Practice, 34*, 353-360. doi: 10.1037/0735-7028.34.4.353

Ward, T., Wilshire, C., & Jackson, L. (2018). The contribution of neuroscience to forensic explanation. *Psychology, Crime & Law*, *24*(3), 195-209. doi.org/10.1080/1068316X.2018.1427746

Weerasekera, P. (1996). *Multiperspective case formulation: A step towards treatment integration*. Malabar, FL: Krieger Publishing Company.

Whitehead, P. R., Ward, T., & Collie, R. M. (2007). Time for a change: Applying the good lives model of rehabilitation to a high-risk violent offender. *International Journal of Offender Therapy and Comparative Criminology, 51*, 578-598. doi/pdf/10.1177/0306624X06296236

Wikström, P. O. H., & Treiber, K. (2017). Beyond risk factors: An analytical approach to crime prevention. In B. Teasdale & M. S. Bradley (Eds.), *Preventing crime and violence* (pp. 73-87). Cham, Switzerland: Springer.

Wilkinson, J. (2005). Evaluating evidence for the effectiveness of the Reasoning and Rehabilitation programme. *The Howard Journal of Criminal Justice, 44*(1), 70-85. doi:10.1111/j.1468-2311.2005.00356.x

Willis, G. M., & Ward, T. (2011). Striving for a good life: The good lives model applied to released child molesters. *Journal of Sexual Aggression, 17*, 290-303. doi: 10.1080/13552600.2010.505349

Willis, G. M., & Ward, T. (2013). The good lives model: Does it work? Preliminary evidence. *What works in offender rehabilitation*, 305-317.

Willis, G. M., Ward, T., & Levenson, J. S. (2014). The Good Lives Model (GLM): An evaluation of GLM operationalization in North American treatment programs. *Sexual Abuse: A Journal of Research & Treatment, 26*, 58-81. doi: 10.1177/1079063213478202

Willis, G. M., Yates, P. M., Gannon, T. A., & Ward, T. (2013). How to integrate the good lives model into treatment programs for sexual offending: An introduction and overview. *Sexual Abuse, 25*(2), 123-142. doi/pdf/10.1177/1079063212452618

Wong, S. C. P., & Gordon, A. (1999–2003). *The violence risk scale*. Saskatoon, SK: Regional Psychiatric Centre and University of Saskatchewan. Retrieved from http://www.psynergy.ca/VRS_VRS-SO.html

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law, 12*(3), 279–309. doi.org/10.1037/10768971.12.3.279

Wooditch, A., Tang, L. L., & Taxman, F. S. (2014). Which criminogenic need changes are most important in promoting desistance from crime and substance use? *Criminal Justice and Behavior, 41*(3), 276–299. doi:10.1177/0093854813503543

Wylie, L. A., & Griffin, H. L. (2013). G-map's application of the good lives model to adolescent males who sexually harm: A case study. *Journal of Sexual Aggression*, *19*(3), 345-356. doi/full/10.1080/13552600.2011.650715

Yang, M., Guo, B., Olver, M. E., Polaschek, D. L., & Wong, S. C. (2017). Assessing associations between changes in risk and subsequent reoffending: an introduction to relevant statistical models. *Criminal Justice and Behavior, 44*(1), 59-84. doi:10.1177/0093854816678648

Yates, P. M., & Kingston, D. A. (2006). The self-regulation model of sexual offending: The relationship between offence pathways and static and dynamic sexual offence risk. *Sexual Abuse: A Journal of Research and Treatment, 18*, 259-270. doi:10.1007/s11194-006-9018-z

Yates P. M., Kingston D. A., Simons D. A., Tyler C. (2009, October). *The good lives model of rehabilitation applied to treatment: Assessment and relationship to treatment progress and compliance.* Paper presented at the 28th Annual Convention of the Association for the Treatment of Sexual Abusers (ATSA), Dallas, TX.

Yates, P. M., Prescott, D. S., & Ward, T. (2010). *Applying the Good Lives and Self-Regulation Models to sex offender treatment: a practical guide for clinicians*. Brandon, VT: Safer Society Press.

Yates, P., & Ward, T. (2008). Good lives, self-regulation, and risk management: An integrated model of sexual offender assessment and treatment. *Sexual Abuse in Australia and New Zealand*, *1*(1), 2-19. ISSN 0833-8488

Zhavi, D. (2006). *Subjectivity and selfhood: Investigating the first-person perspective.* Cambridge, MA: MIT Press.

Ziv, R. (2018). *The future of correctional rehabilitation: moving beyond the RNR model and good lives model debate*. New York, NY: Routledge.