Cyberbystanders:
Behind the Screens

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Psychology

at

The University of Waikato

by

Emma-Leigh Ann Gilmore Hodge

KO TE TANGATA

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2023

Dedicated to Kitty Genovese,

And to any disempowered person whose story has been minimised rather than supported, with all its beautiful complexities remembered.

Abstract

Behaviour that occurs online is an important component of many people's social interactions. It is, therefore, crucial that the detrimental impacts of cyberbullying and similar behaviour (i.e., online aversive behaviour) are addressed. To this end, I present research that progresses from an exploration of online social behaviour in general (Chapter 2) through to quantitative discounting studies that examine an avenue for behaviour change initiatives that improve the online social environment.

In the exploratory phase (Chapter 2), I conducted a series of focus groups with young people between 12- and 21-years old on the broad topic of Social Networking Site use (SNS). Participants spoke about how they manage their use of SNS, which coalesced around several key themes. One of these ideas was the importance of showing and receiving support from their online social communities, especially in instances of adversity.

An effective way to show support for another person online is to actively intervene when witnessing adverse (or aversive) behaviour directed toward that person. In Chapter 3, I explored a novel application of discounting methods to assess whether social outcomes related to online bystander (i.e., cyberbystander) intervention are discounted similarly to outcomes more common within the discounting literature. I used discounting methods in conjunction with the framework of the Bystander Intervention Model (BIM; Latané & Darley, 1970), reframing the latter to be congruent with the concepts of behaviour analysis.

In Chapter 4, I tested the applicability of discounting methods to the decision-making of young adult cyberbystanders. As discounting studies using monetary outcomes have been well-established, I compared the probabilistic discounting of money to that of social outcomes related to cyberbystander behaviour. Having established that probability discounting can be applied to cyberbystander's decision-making, in Chapter 5, I examined whether scenario severity, audience size, or locus of responsibility (i.e., whether responses related to

participant's own behaviour or their perception of how someone else would respond) impacted the willingness of young adults to intervene in online aversive behaviour. Intervention was marginally more likely when participants considered whether someone else would intervene (rather than themselves) and if the severity were high, but the larger audience size had a negligible effect. Chapter 6 contains a study designed to follow up on the promising findings regarding the impact that perceived scenario severity had on cyberbystander's decisions to intervene themselves, particularly when compared to what participants considered the social norm to be (i.e., how they thought someone else would respond). The difference between the likelihood of intervention when participants responded for themselves versus what they considered normative was most evident when the scenario was moderate-to-high in severity.

In all, my research supports the premise that discounting methods can be applied to cyberbystander decision-making and used to assess aspects of the BIM to identify avenues for cyberbystander behaviour change initiatives. Cyberbystanders' actions can fundamentally change whether acts of aversive behaviour are reinforced (or punished), as well as mitigate the extent of the harm experienced by a target of online aversive behaviour.

Acknowledgments

Thank you to everyone who ever encouraged me to carry around a notebook, to ask questions, and to at least start things. I have not historically been good at finishing projects, but this chapter (well, seven chapters) is now complete, and I could not have done it without plenty of help.

Thank you to my supervisors who guided me through this – Dr. Neville Robertson, Dr. Tim Edwards, and, of course, Dr. Rebecca Sargisson. Rebecca, you have easily impacted me the most throughout this journey – professionally, academically, and personally. Professionally, you gave me my first shot at teaching and kindly pointed out, "The students already know you're a geek, so you may as well lean into it" (or something like that), which was incredibly freeing. Academically, finishing this thesis would not have been possible without your support and guidance. Your encouragement was so great that, for a while, I had forgotten that the initial idea to research online behaviour had not been my own – it has been a privilege to pick up that ball and run with it. Personally… I am sure I had (now lost) a file of wonderfully written thank you's on this personal note – complete with plenty of hyphens. Perhaps not, though. Perhaps I never had the right words to feel that I had thanked you enough. After all, you got me back watching Survivor! Thank you for everything, R.

Thank you to Emma Rockel, who has seen me at my worst more than anyone (correlation is definitely not causation!) but has always been there for me. Through my studies, you have reliably been unsurprised that I was doing well, which has been of huge help in keeping my own doubts at bay.

Thank you to all I have met through my studies – fellow students, colleagues, and those who have been both. A few in particular: Juliana Brown, who was of enormous support in the most challenging year of this process; Tina Bourke, who has always reminded me there is "only one way to eat an elephant"; Tegan Andrews, who spent many hours listening to my bewilderment over methods minutiae; Justin "just get it done" Phillips who (unknowingly, I think) helped me refocus when I could have gone down an especially large rabbit hole; and, Tracey Morgan, who invited me into the fold with CeTTL which considerably reduced the number of non-thesis things I had to stress about.

And to the many others I have not been able to list, my parents, sister, family, and friends who all played their important parts, thank you!

Table of Contents

List of Figures

List of Tables

List of Terms

Abbreviation	In-Full
SNS	Social Networking Sites, encompassing social media (SM) sites
BIM	Bystander Intervention Model, as proposed by Latané & Darley (1970)
MCH	Monetary Consequence, High Severity study condition
MCL	Monetary Consequence, Low Severity study condition
ECH	Emotional Consequence, High Severity study condition
ECL	Emotional Consequence, Low Severity study condition
VAS	Visual Analogue Scale, as used in Kaplan et al., 2014
TO	Time Online; self-reported ratio measure
ToSM	Time on social media (i.e., SNS); self-reported ratio measure
PSES	Perceived Socio-economic Status; modelled off Adler et al., 2000
RSE	Rosenberg Self-Esteem Scale, from Rosenberg, 1979
P.ECH	Personal responsibility to intervene, high-severity scenario
O.ECH	Someone else’s responsibility to intervene (i.e., an “other” person), high-severity scenario
P.ECL	Personal responsibility to intervene, low-severity scenario
O.ECL	Someone else’s responsibility to intervene (i.e., an “other” person), low-severity scenario
PL	Personal responsibility to intervene, large-audience scenario
OL	Someone else’s responsibility to intervene, large-audience scenario
PS	Personal responsibility to intervene, small-audience scenario
OS	Someone else’s responsibility to intervene, small-audience scenario
SOAB scale	Severity of online aversive behaviour scale

Chapter 1: General Introduction

Part 1: Online Social Behaviour

The internet is a critical part of many people's daily lives, especially in terms of staying socially connected (Cassidy et al., 2013; Heflin et al., 2017). For instance, use of the internet and social networking sites (SNS) can help maintain long-distance friendships (e.g., O'Reilly, 2020), establish new relationships (e.g., Burke et al., 2011; Lauricella et al., 2014), pursue professional opportunities (e.g., Utz & Breuer, 2016), and overall can increase an individual's social capital (i.e., social resources they have available; Kraut et al., 2002; Williams, 2006). Additionally, use of SNS may be particularly important for those who are most susceptible to marginalisation. Put another way, ready access to SNS and the online communities that utilise them can be a protective factor for groups that would otherwise be vulnerable (Mueller, 2019).

For some researchers, the ubiquity of SNS usage is of concern because there may be negative impacts on facets of mental well-being (e.g., Cunningham et al., 2021; Moreno & Whitehill, 2014; Twenge et al., 2018). While the research literature contains mixed information regarding whether general SNS use is a net-good, there is clear evidence that cyberbullying and similar types of behaviour are prevalent (Kowalski et al., 2014; Kowalski et al., 2019) and have detrimental consequences (Lysenstøen et al., 2021; Zhu et al., 2021). In this thesis, I employ the term "online aversive behaviour" to encompass all actions causing perceived harm to others, including but not limited to cyberbullying (Byrne et al., 2016; Vaillancourt et al., 2008).

Becoming a potential target of online aversive behaviour is a pertinent social issue due to the integral role that SNS plays in facilitating social connectedness (Marciano et al., 2022). Online aversive behaviour is sometimes characterised as the result of specific technological platforms (e.g., SNS) or advancements (e.g., smartphones or video messaging; Catone et al., 2020; Sung-min & Byong-jin, 2022). However, the consistent underlying factors are the actions of people and so a science of human behaviour is critical for understanding and addressing harmful online behaviour (Skinner, 1971).

At the outset of my thesis research, it was clear that an initial exploratory phase was required to narrow the subsequent research focus to a specific aspect of social behaviour online. Therefore, I began with a series of exploratory and semi-structured focus groups with adolescents, as it is adolescents who are most impacted by cyberbullying and similar behaviours (Cabello-Hutt at al., 2017; Kowalski et al., 2014; Shapka et al., 2018). Using examples from focus group participants, I overviewed the general topic of online social behaviour in Chapter 2. I identified key themes of interest, including identifying aversive behaviour and the actions of cyberbystanders as a focal point for later phases of this thesis.

Participants in my focus groups addressed their concerns and spoke of various harmful impacts related to social networking sites (SNS). Moreso, they gave me an overarching view of how SNS fit in with and augmented their everyday behaviour – in good and challenging ways. They shared insights into their strategies for SNS use, which revolved around several central themes that I present in Chapter 2. The significance of seeking, offering, and receiving support in online social spaces emerged as a key theme from participants. I, therefore, focussed the later part of this thesis on the behaviour of cyberbystanders, who are well-placed to provide support and mediate the impact of aversive (i.e., harmful) behaviour conducted online (Barlińska et al., 2015; Macháčková et al., 2015; Zhao et al., 2023).

Part 2: Cyberbystanders

Cyberbystanders have the potential to play a pivotal role in what behaviour is and is not accepted in online spaces (Allison & Bussey, 2016; Macháčková & Pfetsch, 2016). For instigators of aversive behaviour such as bullying (i.e., aggressors), their bullying behaviour is often controlled – behaviourally speaking – by social reinforcement provided by bystanders

(Guerin, 1994; Safin et al., 2015; Salmivalli, 2002). Research concerning bystander behaviour has a long history (Latané & Darley, 1968) and has been examined in relation to different kinds of behaviour, including medical events (e.g., Latané & Dabbs, 1975), sexual harassment in the workplace (e.g., McDonald et al., 2016), and in response to relational violence (e.g., Burn, 2009; McMahon & Dick, 2011).

Several phenomena that might either increase or decrease the likelihood of bystanders intervening have also been identified. For example, early studies into the bystander effect found that the more ambiguous a scenario is, the less likely it is that bystanders would intervene (Latané & Darley, 1970; Latané & Rodin, 1969). Factors contributing to ambiguity include anonymity, which can obscure the identities and relationships of the individuals involved (Solomon et al., 1982). Additionally, scenarios could be situationally ambiguous when bystanders struggle to differentiate between genuine incidents and performative stunts (Phares & Wilson, 1972). It is also possible that intervening in online aversive behaviour exposes cyberbystanders to risks, counteracting the principle of self-preservation (Abbate et al., 2022). Put another way, intervening to help someone else could instead make them the target of further aversive behaviour; thus, intervening might function as a positive punisher that decreases the likelihood of cyberbystander intervention (Baum, 2017).

The interplay between self-esteem and online behaviour also warranted attention because being bullied often correlates with lower self-esteem, reduced social capital, and increased isolation (Burger & Bachmann, 2021; Cramer, 1990; Kearney & Silverman, 1995; Lei et al., 2020). In regard to bystanders, researchers have linked self-esteem and intervening to help others in that those with higher self-esteem and self-efficacy are more likely to intervene (McCarty et al., 2016; Paull et al., 2012; Sheeran et al., 2014). Self-esteem and other indicators of well-being may also be indirectly linked to cyberbystander behaviour through time spent online (Kross et al., 2013; Quiroz & Mickelson, 2021). Some researchers have found a correlation between lower levels of self-esteem and high rates of internet use (Twenge, 2017; Twenge et al., 2018; Vannucci et al., 2019) and spending more time online increases the chance of being a cyberbystander to aversive behaviour (Park et al., 2014). Moreover, there is already a growing body of research into how bystander behaviour functions in online spaces. For example, the work of Ferreira et al. (2020) on cyberbystander behaviour and self-efficacy, as well as Macháčková et al. (2015) and their work regarding audience size and cyberbystander behaviour.

Part 3: Model & Methods

The Bystander Intervention Model

Much of the research about bystander behaviour utilises the Bystander Intervention Model (BIM; Latané & Darley, 1970). A benefit of using the BIM to frame cyberbystander decision-making is that it allows for the study of each link (i.e., the steps) in the chain of contingent behaviours that culminate in the act of intervening. This is important because the factors that influence cyberbystander decisions are likely to be complex (Allison & Bussey, 2016; Kowalski et al., 2014), and so determining which factors interact with which steps should contribute to more effective behaviour change initiatives. Although there is a substantial body of literature in which researchers have applied the BIM (see Fischer et al., 2011), few include an empirical test of the proportion of people who fulfil the early steps in the model (Mennicke et al., 2022). It is useful to measure how many people are likely to perform an intervention (i.e., Step 5), but as a step-wise model, it is also critically important to measure how many people reach prior steps (e.g., Steps 2 & 3) and to investigate any barriers to progressing to the later steps. As an additional component of this thesis, I tested the proportion of participants who fulfilled Steps 2 and 3 of the BIM and whether doing was related to factors such as the scenario severity.

Discounting

Chapter 3 introduces the idea of applying a method used in behaviour analysis – discounting – to measure cyberbystander's intentions to intervene to help targets of online aversive behaviour. Moreover, I used discounting methods to better understand the environmental factors that increase or disproportionately decrease (i.e., steeply discount) cyberbystander's willingness to intervene. Discounting methods have been useful for researching risk-laden decisions such as smoking (Reynolds, 2004), substance use (Mejía-Cruz et al., 2016), becoming distracted while driving (Hayashi et al., 2016), and in cases where the behaviour could be environmentally costly (Carson & Tran, 2009; Sargisson & Schöner, 2020). Chapter 4 contains my Pilot Study (4.1), in which I compared probabilistic (i.e., the level of ambiguity) social outcomes that are related to common cyberbystander scenarios against monetary outcomes, for which discounting methods have more extensively been applied. Then, having demonstrated the applicability of discounting methods to cyberbystander decision-making, I investigated the effect of differing severity levels, audience sizes, and whether participants would take personal responsibility for intervening (Chapter 5).

I found that a participant's perception of the severity of a scenario is of utmost importance in deciding about whether to intervene. Scenario severity may interact with fulfilling Step 2 of the BIM: the more severe a scenario is, the more likely it is that participants will recognise that intervention is required. Building upon the design of my probability discounting studies (i.e., Studies 4.1, 5.1, & 5.2), I developed a severity discounting procedure to examine further how severity impacts cyberbystander behaviour, which I present in Chapter 6.

Summary

In sum, the ideas from the bystander effect literature have been applied to the understanding of comparable social issues (e.g., Burn, 2009; Fischer et al., 2011; Jönsson, & Muhonen, 2022), and for this reason, I applied these ideas to cyberbystander decision-making. Cyberbystanders are, by definition, more removed from an instance of aversive behaviour than either targets or aggressors and they are, therefore, prime candidates for behaviour change initiatives because their future behaviour is less restricted by pre-existing reinforcement histories (Cassidy et al., 2013). In contrast, it tends to be more difficult to guide an aggressor to change because their future behaviour is likely to be biased toward carrying out the aversive behaviour already embarked on.

Each quantitative study presented in this thesis supported the premise that discounting methods could be of use in understanding the drivers of cyberbystander decision-making, including how these drivers interact with the steps of the BIM. Furthermore, others' tests of the BIM and its constituent parts have been used to develop behaviour change initiatives in offline contexts (e.g., Lassiter et al., 2021; McDonald et al., 2016; Worsteling & Keating, 2022), and my thesis provides support for the applicability of the BIM in developing cyberbystander behaviour change initiatives for mediating the impact of aversive online behaviour.

Chapter 2: Focus Group Research

Outline

In this chapter, I present the findings from my qualitative work in conjunction with an overview of research literature that concerns online social behaviour. I facilitated focus groups with a range of young people to discuss, primarily, the social networking sites (SNS) they used and how they and their peers utilised these for connecting socially. Focus group participants reported using a wide variety of SNS, but there was considerable overlap in how different platforms were used. Put another way, differences in the form used for the social behaviour (e.g., a specific SNS) were less important than the function (e.g., wanting the communicate one-on-one). Participants gave an overarching view of how SNS fit in with and augmented their everyday behaviour – in good and challenging ways. Participants also recognised that anonymity, the potential for misunderstanding what others intended (i.e., behaviour being ambiguous), and the ease with which information spreads are all factors that can (in some cases) make the online social environment more volatile than the offline one. In response to this recognition, participants generally showed self-awareness and provided their ideas for improving the experience of online social environments.

Across the nineties (1990-99), internet access went from roughly 14% to 45% of U.S. adults (Kraut et al., 1998; Pew Research Center, 2014), although estimates for some groups were considerably higher (e.g., approximately 80% of university students, Anderson, 2001). The quick rise in internet availability began impacting Aōtearoa New Zealand households in the mid-nineties (Brown, 2014) – my household included. From the outset of my doctoral studies, I planned to carry out focus groups with people ranging from approximately 12 years old to 20 years old to investigate how social behaviour functions in online spaces. Adolescents and young adults are likely to be most impacted by changes in the technologies that augment social behaviour, such as the internet initially and now also SNS (e.g., Livingstone et al., 2017; Polanco-Levican & Salvo-Garrido, 2021). However, as the facilitator of exploratory focus groups, it was crucial that I was aware of my relevant biases in regard to online behaviour. It was equally crucial that I was aware of my biases during phases of analysis and, moreover, that my pre-existing history with the topic was presented transparently alongside a discussion of those analyses (Howitt, 2019; Morgan, 1996).

My Approach to Focus Groups

For me, childhood internet use was dominated by the abrasive dial-up tone, flash games, and IRC-style (i.e., Internet Relay Chat) chat rooms. Then came the instant messaging platforms from the likes of MSN and AOL (Gross et al., 2002), which are functionally not all that different from present-day instant messaging applications. Around this time, socialising platforms incorporating elements of gaming (or games incorporating social aspects) were gaining popularity. For example, Halo and World of Warcraft were released in the early 2000s, each with online multiplayer options. Several major gaming consoles also integrated internet access; a feature that would become standard.

To add to my own experiences growing up in the increasingly connected world are the many others born since the 1990s well-represented in the Social Networking Site (SNS) research literature. This group born in or since 1990 is often referred to in the collective as "digital natives" (Prensky, 2001). Digital natives are in an interesting position in that they are the experts in how online communication occurs. The reason is threefold: younger people are most likely to interact with the newest advances in social networking technologies (Wartella et al., 2016); then for adolescents are the most likely to explore the social opportunities available to them (Lauricella et al., 2014), and to have the independence to make their own decisions on how to do so (Fernández Morales, 2021). The internet has many clear applications, many of them productive and positive (e.g., Eppard et al., 2016; Heflin et al., 2017), though there must also be caveats to its use (Seo et al., 2016). Therefore, my starting point for this thesis was a series of wide-ranging discussions with adolescents. My focus on adolescents is not because I believe they are the only group affected but because they are at the forefront of any social issues related to the growing ubiquity of online communication (Bauman & Belmore, 2015; Prensky, 2001).

Online Social Behaviour

Some researchers have suggested that one in three adolescents prefers to have deeply personal conversations over a device, which some take as indicative of adolescents' failing to communicate in-person due to an overreliance on devices and the internet (Schouten et al., 2007). Similarly, Kielser et al. (1994) are often cited as a source of the supposition that social interactions conducted in online spaces are, by nature, much more likely to engender conflict (e.g., Ho & McLeod, 2008; Pyžalski et al., 2022). Kielser et al.'s initial work on the topic was explicitly stated to be speculative and intended to generate future research – not as a factual account of how socialising on the internet functions (Matias, 2017). Furthermore, findings from Sproull and Kiesler (1986) at around the same time were null or inconclusive regarding whether online spaces impacted negatively upon the valence of people's social interactions.

Nevertheless, it was their earlier (i.e., Kiesler et al., 1984) research that showed that the anonymity of the internet *might* create more aggressive interactions that has endured the most.

There is contrasting literature regarding the benefits of socialising using SNS, albeit a smaller amount. For example, if people are (or move) further away from each other and utilise remote working and studying options, digital technology provides greater opportunities for connectedness (Wang et al., 2014). In addition, groups and relationships formed around specific interests (e.g., particular games) or characteristics (e.g., members of LGBTQIA+ communities) are opportunities for connection that are more readily available to today's young people (O'Reilly, 2019). That is, groups of like-minded people - which previously only the lucky stumbled upon – are much easier to seek out and find.

The content and the accessibility of the internet have increased exponentially since its beginning in the 1980s (Castells, 2002), not least of all in the wealth of opportunities to generate social content and have a social "presence" online. Despite the growth of the internet its basic structure has remained largely the same - open with few boundaries. That said, a substantial change in how today's internet operates is that much social communication is filtered through a relatively small number of Social Networking Sites (SNS). Furthermore, many of the SNS available are now owned by an even smaller number of businesses with goals that may not align with their users' (Banchik, 2020; Sagioglou & Greitemeyer, 2014; Tromholt et al., 2015). For instance, an algorithm that was central to the workings of a particular SNS was changed without notification in 2018 to promote "meaningful interactions". Not long after implementation, the outcome of reinforcing content that was especially engaging became clear: the SNS was now rewarding and promoting controversy, extremity, and outrage. Although evidence of the detrimental effect this algorithmic change caused was being generated daily, little organisational attention was given to the issue until internal documents were made public in 2020 (Gopal, 2021; Hagey & Horwitz, 2021).

While substantial technological advancements are being developed and applied to SNS (e.g., natural language processing, personalised advertising algorithms), there has been comparatively less attention paid to the impact that these changes have on human behaviour. The substantive gap between the number of SNS employees whose roles involve the regulation of harmful content and those whose involve innovation of their platforms is congruent with businesses prioritising technological advancement over the behavioural impacts on users of SNS (Banchik, 2020). Skinner (1971) noted that there are many branches of science that have been advanced much further than the science of human behaviour has been – to the detriment of society. "Physics and biology have come a long way, but there has been no comparable development of anything like a science of human behavior," Skinner (1971, p.5) asserted. Computer science and all the technical inputs required to run SNS are another example of a field that has been advanced ahead of an analysis of behaviour. SNS businesses should be held accountable for their part in the proliferation of harmful online behaviour. However, the focus of my research was on behavioural explanations of cyberbystander intervention and its role in reducing the impact (and amount) of harmful behaviour on SNS.

Current Research

For this first study in my thesis, I conducted a series of focus groups to examine the topic of online social behaviour from the viewpoints of those most immersed in it. I had broad, exploratory aims for this study but hoped to pose an answer to the following research questions:

RQ2.1: What SNS do young people use, for how long (time spent on social media), and in what ways (i.e., for what function)?

RQ2.2: What are the experiences of young people who use the internet to socialise and how do these experiences compare to extant research literature that documents harmful online behaviour?

RQ2.3: What kinds of support do young people have around them in regards to their online behaviour, and, in which cases is that support most useful (or a hinderance) to them?

Ultimately, I was interested in better understanding young people's use of SNS so that I could later investigate avenues for lessening the prevalence and impact of harmful online behaviour. It is worth caveating that whatever those avenues may be, it is unlikely that they will be straightforward or complete solutions. That said, a central premise of this thesis was that the pursuit of – if not the actual attainment of – a less combative and overall, more inclusive online social environment is a necessary goal.

Methods

Participants

Recruitment

Over the course of 2016 and 2017, I conducted seven semi-structured focus groups to explore young people's behaviours and opinions about communicating via social media sites. Six focus groups were organised with the help of the participants' respective school personnel, which allowed me to use convenience sampling to ensure a broad range of individuals were included while also incorporating peer groups (Howitt, 2019, p. 88; see Appendix A). Included schools were predominantly located in Tauranga, New Zealand, although two (which will be noted) were in the nearby city of Hamilton. The participants of the seventh focus group all frequented the same non-governmental organisation in Tauranga, and so a similar convenience sample was possible, albeit with organisation administrators rather than teachers. The relevant organisers, teachers or administrators distributed study signup information and consent forms without the presence of our research team, ensuring participation was voluntary. Only those with an interest in taking part and who had returned completed consent forms in time for the scheduled focus group participated. My approach to conducting the focus groups was intentionally broad and exploratory in nature, intended to gather a range of views and personal experiences from young people (Morgan, 1996).

Most of the extant literature has been written about participants in high school (Bauman & Bellmore, 2015), making it difficult to explore potential age differences. While many social networking sites (SNS) administrative teams attempt to limit access to people under 13 years old, it is widely reported that age guidelines are easily bypassed (Blackwell et al., 2014). Relatedly, there is an increasing number of individuals in this age group who own a personal cell phone but are theoretically constrained by these SNS age restrictions, as well as more stringent school rules (e.g., Hodge & Sargisson, 2017), which make their experiences and online behaviour of particular interest. That is, these under-aged (for SNS) participant experiences may indicate whether the current attempts to restrict their internet use are effective. Therefore, I included participants who were 12 to 20 years old.

Focus Groupings

Two focus groups were with pre-adolescent students who attended classes with a particular focus on incorporating technology use in their education. At the group level, I refer to these as P-A1 (Pre-Adolescent Group 1; $n = 3$) and P-A2 (Pre-Adolescent Group 2; $n = 4$). Other than a slight difference in numbers and a reversed gender skew where P-A1 was majority boys, and P-A2 majority girls, these groups were very comparable in their make-up.

One focus group involved three participants who all attended the Junior years (i.e., 13- to 15-years old) of a local all-girls high school, referred to subsequently as A-G (Adolescent, Girls School).

Located in a nearby city (Hamilton), I also carried out a focus group with six high school students, for whom there was some age overlap (14- and 15-year-olds) with the previous grouping, though this focus group involved participants of mixed gender and will be noted as A-M (Adolescent, Mixed Gender).

Seven young people of mixed gender (approximately 17 years old) attending a local Catholic secondary school participated in focus group T-C (Tauranga-Catholic), and by way of rough comparison, another group included six young men attending their final year of a “single-sex” Catholic school in Hamilton (H-C; Hamilton-Catholic).

A further six participants local to Tauranga took part in the organisation-based focus group and were between 18- and 20 years old at the time of discussion, which I refer to as O-B (organisation-based) hereafter.

I introduced the focus groupings largely in age order but also to indicate some key sets of comparisons. For example, the two focus groups where sets of attendees were from one of two Catholic schools located in different but broadly comparable New Zealand cities. Focus groups A-M and H-C can be considered comparable in some important ways too. Both sets of participants attended schools in the same Hamilton neighbourhood and are close to each other, which may mean overlap in social circles beyond the school gate. That said, the A-M participants were likely to be moderately lower in socioeconomic status, according to the school’s comparatively lower decile rating (Ministry of Education, n.d.). Further, the A-M participants’ school was secular, unlike the explicitly Catholic establishment the H-C students attended.

Table ***1*** contains an overview of the similarities and contrasts between the focus groupings.

Table 1

Matrix of Focus Grouping Characteristics

Code	Age Range	No. of Participants	P-A1	P-A2	A-G	A-M	T-C	H-C	O-B
P-A1	11-12	3 (1)		*Different Class	Age	Age	Age	Age	Age
P-A2	11-12	4 (3)	*Otherwise very similar		Age	Age	Age	Age	Age
A-G	13-15	3 (3)	City	City		City	Age & religion	Age, religion, & city	Age
A-M	14-16	6 (3)			Age		Age, religion, & city	Age & religion	Age & city
T-C	17-18	7 (3)	City	City	City			City	Age
H-C	17-18	6 (0)				City	Age & religion		Age & city
O-B	18-21	6 (3)	City	City	City		City		

Notes. Participant numbers are presented alongside (in brackets) the number of those who identified as female. Codes for each focus grouping are listed (top-to-bottom, left-to-right) in the order they are introduced in-text. In the top triangle (shaded red) are noted the relevant contrasts between each pair of focus groupings (e.g., the A-G and P-A1 groups differ in age); the lower triangle (shaded green) show the similarities across the groupings (e.g., the A-M and H-C groups were both located in Hamilton). The astrix indicates two groups that are particularly similar as they are from the same school and are the same age.

Procedure

I arranged discussions so that each group consisted of three to seven participants to balance the competing needs of appropriate facilitation and free flowing (i.e., independent of facilitator input) conversations (Howitt, 2019). I first asked that participants introduced themselves and named the SNS that they had used the most in the past week. I followed this by lead a discussion about any other SNS that they used over the course of an average week, which provided opportunities for discussing the commonalities and distinctions between their use of different platforms. Oftentimes, these broad initial questions were sufficient and provided me with plenty of avenues to follow-up. All focus groups were semi-structured, particularly as I hoped to hear candid responses of the kind that teachers and guardians may not be readily aware. At a point when the freeness of conversation lessened or when my time was nearly up, whichever came first, I prompted a conversation about the valence of several verbatim SNS comments. As well as more general comments, I asked that participants give each of the comments a severity rating between 1 (very negative) and 5 (very positive), with 3 as a neutral point. I used the severity rating exercise as a tool to guide participants into a discussion about how they interpret and react to different types of online social behaviour, and I selected the comments specifically with this goal in mind (see Appendix A for further information regarding the focus groups).

All focus groups were video and audio recorded with participants' consent and ranged from 45- to 50-minutes in length. I then prepared full transcripts of the conservations. Through the transcription process, I coded the individual participant's contributions to retain general, non-identifiable information about each of them (e.g., their age). My focus group research gained ethical approval from the University of Waikato's School of Psychology committee (Protocol #16:53).

Analysis

I analysed the information collected across the seven focus groups using a mix of qualitative and quantitative methods.

Primarily, I conducted a thematic analysis (Howitt, 2019) and made use of verbatim quotes to represent the themes while still retaining individual participant voices. My analysis of the transcripts was iterative, from the transcribing phase through several stages of review, continuing into the draft writing stage (Morgan, 1996). To complement these themes, I collected quantitative information about which SNS participants used, estimates of their time spent online, and the severity ratings given to each of the 10 SNS comments.

Findings & Discussion

In the early stages of each discussion, I established that there were three SNS that most young people used over an average week; these were Instagram, Facebook, and Snapchat. Oftentimes Facebook Messenger was also mentioned as a separate entity from just "Facebook", though I collapsed both into one category as they are widely seen as indistinguishable, at least at a platform level (O-B & T-C).

> O-B/ A. Can you even use Facebook without...like, if you want to talk to people, can you even use Facebook? Coz it made me download Messenger to talk to people. I don't even think on Facebook you can...you can't just download Facebook and talk to people you have to download Messenger at the same time. Unless you're on like a laptop or something like that.
>
> T-C/ A. But—but in Facebook...I personally include that as Messenger. It's a separate app but…
>
> Yeah, it is *Facebook* Messenger [emphasis the participant's own]

Once we had covered the initial, most-used SNS, I encouraged participants to consider a broad definition of what counts as an SNS (Verduyn et al., 2017), which participants readily incorporated into their conversations. A participant in the A-M focus group summarised what came to be my working definition of what social media is (i.e., SNS):

> A-M/ I guess anything that…that you can leave a comment on, or you can talk to other people on, is considered social media.

In Table 2, I present the frequency with which each subset (i.e., each focus group) mentioned using a particular SNS. The three most used SNS appear in the first three columns, whereas the fourth contains "Musically", included separately because of the platform's subsequent evolution into TikTok.

Table 2

SNS Used by each Focus Group

FG#	*n*	Instagram	Facebook	Snapchat	Musically
P-A1	3	**3**	**3**	2	2
P-A2	4	3	2	3	3
A-G	3	2	**3**	1	2
A-M	6	**6**	**6**	**6**	
T-C	7	4	5	6	
H-C	6		5	5	
O-B	6	**6**	**6**	**6**	
N =	35	14	14	12	7
		40%	40%	34%	20%

The results in Table 2 are not exhaustive but do represent the SNS that participants used most often. Participants mentioned a wide variety of other SNS-type platforms, including games that had a social component (A-M & A-G).

> A-M/ You have ones [SNS] that people don't really think of as communication…like…you could go to the point where you say Xbox. You know…it is all communication. I could get onto my Xbox and talk to people so that's communication.

A-G/ A. Yeah, most games do [have a chat component, like QuizUp] coz that's what allows people to talk more, make friends...

Having established which SNS participants used, I then guided the discussions toward how participants used these SNS. The only apparent platform-based difference was the tendency of different age groups to prefer one over the other (A-G & O-B). Any preferences of core SNS were more related to the restrictions or trends most prevalent for that group rather than any personal preference (e.g., T-C).

A-G/"...[I] don't really find the difference between them all [SNS]".

O-B/ Oh yeah, I just use all of them. I don't…most people – oh, if I'm not wrong – would probably use most of the websites. …all of them pretty much.

T-C/ A. My sister…like…Instagram's her main Messenger

B. Yeah! I've noticed a lot of younger kids, coz they're not allowed Facebook, they use their um…source of like— like direct messaging?

A. I think that's because parents don't like the whole...Facebook thing. But they think Instagram...I don't know…

Furthermore, in two cases, participants began explaining that the three core SNS were very different from one another but ultimately decided that this was not the case (e.g., O-B & P-A2). In these cases, the function of the behaviour was more important than which specific platform the behaviour occurred on. For example, a participant wanting to send a photo to a select group of friends (i.e., the function; Haynes & O'Brien, 2000) was more relevant than whether they chose to do this on Instagram versus Facebook (i.e., the form).

O-B/ A. Yeah coz...yeah, Snapchat and Facebook are quite different.

Researcher. Yeah sure Snapchat for photos, Facebook for text...?

B. Ohhhh...[sounding unsure]

A. Oh well, I mean...you can...for Snapchat you can do both. Like you can text or can send photos

P-A2/ A. Yeah, I think Facebook and Instagram are really similar, except –

B. ...exactly the same.

The various core platforms all appear to provide participants with an entry point to a conversation with someone they know in similar ways. The passages from focus groups O-B, P-A2, and H-C below exemplify this idea best. Further, Instagram and SnapChat were both referenced by participants as good vehicles for following celebrities they have an interest in. Specific examples were in the T-C and the H-C focus groups. Several groups also referred to having spaces set up online for them and their classmates to communicate primarily, but not exclusively, for study purposes. Even within one focus group (A-G), where participants were from the same school but did not share classes, both Instagram and Facebook were mentioned for separate within-class group chats.

O-B/ Coz I guess on Snapchat people put up photos of what they're doing throughout the day…and then you can just, comment on that photo and it gets sent straight to them so you can start talking about that, then just carry on from there.

P-A2/ I use it [Instagram] mainly for like, yeah, seeing what people are up to and...chatting to people.

H-C/ I think Facebook is more having that access to information and then being able to tag your friends... I'll tag this friend in the comments so they can...they can catch up on it and we can have a discussion

T-C/ Yeah I think something I use Instagram for is...also, like um, keeping up to date with like celebrities or sports personalities that I like. Seeing what, like, idols are up to and stuff—

H-C/ A. Yeah…I follow some like Basketball teams and stuff [on SnapChat]

B. Yeah like, I follow some of the NBA teams and it shows like their practises and like has…casual interviews with the players.

Regarding Musically, Table 2 highlights that only the younger participants from the pre-adolescent (P-A1 and P-A2) and the A-G focus groups thought to mention it as an SNS. Given the subsequent rise of the platform, largely after its evolution into TikTok, this may reflect a change in popularity trends and that these are often paired with particular age cohorts (Pew Research Center, 2021). Although most independent data about SNS usage relate to advertising opportunities and, as such, only captures 18+-year-old usage, there are clear age-based differences in platform use, especially for TikTok and SnapChat. The 18- to 29-year-old group reported using both SNS (TikTok = 48%; SnapChat = 65%) more often than the next age group (20- to 49-year-olds, 22% and 24%, respectively; Pew Research Center, 2021). Also, noteworthy in the Pew Research Center (2021) data is that use of Facebook was the least variable across the whole range of age groups questioned (18-29 up to 65+), which was supported by comments in the O-B and H-C focus groups. As the data reviewed here shows *use of* rather than the *frequency* of use, Facebook is the most consistently used across the age spectrum.

O-B/ I talk to grandparents and like – older cousins on Facebook, whereas I talk to people my own age on SnapChat. Just coz it's a little bit easier to understand for older people.

H-C/ Just use it [Facebook] for sort of looking at news and for looking what friends are up to and...communicating with friends and family and...yeah.

Beyond discussions of the core SNS and of Musically, participants used a wide range of other social media platforms. Three SNS had a heavy focus on their user's anonymity (YikYak, Sahara, and Ask FM) and I categorised these under the theme *anonymous*. I found that only participants who were 15 years old or above mentioned wholly anonymous platforms. While any type of communication has the possibility of carrying positive or entirely neutral valence, which was carefully acknowledged by participants (e.g., A-M quote below), discussion of behaviour on these anonymous platforms tended toward being harmful:

> T-C/ It [YikYak] was just really just like...like, anonymously bully people
>
> A-M/ A. Yep and [Sahara's] become a big thing…that...it can be helpful...it can be constructive, like, I have seen a lot that are just like "You're beautiful, you're amazing"... [But] then there's really like...obscene, gross stuff. Like stuff that should really just not be seen...like...it's become really sexual also.
>
> B. Yeah stuff that shouldn't be said. Period.
>
> T-C/ Like…[Ask FM's] meant to be a thing where you confess your crush but people just go on and like...they…they're brutal!

In all cases, these platforms had clearly been designed to incorporate the highest level of anonymity while also allowing for the anonymous threads of conversation to be shared by the recipient on other SNS. In at least one case, the anonymous platform was directly integrated into one of the core SNS, so a person's online contact could migrate to this sub-platform if they wanted to comment incognito. This feature increases the likelihood of harm due to the implication that any cruel messages left for the user are, in fact, from someone they know personally, perhaps even someone they believe to be a friend (Sticca & Perren, 2013). Furthermore, the level of anonymity that online perpetrators perceive they hold has been established as a meaningful predictor of subsequent cyberbullying behaviour (Barlett et al., 2021), as has having a high belief in the impermanence and the ability to erase content online (Wright, 2013). The anonymous platforms mentioned by my participants appear to have both these features:

> T-C/ A. Yeah...you leave a "constructive" message [laughing from others] and what people have been doing...and I did it...it was um, you put the link on your SnapChat

Story so people open the link and you just literally just leave a message and...coz heaps of my friends did it so like...peer pressure [laughing a bit]. Not peer pressure really...but…

B. Positive peer pressure?

A. [I] did it [use Sahara] and yeah...I'm not going to do it again I just did it for LOLs

<u>Researcher. Yeah...so how positive does that kind of thing go?</u>

A. Some people would—

C. --It's pretty dodgy!

B. Not positive [general agreement here]

A. Some people would use it—

C. It's a pretty...dangerous position for yourself

A. --to hurt people

As well as anonymous platforms heightening the possibility of receiving harmful messages online, my participants shared complex views on the fact that users first need to engage with this kind of platform to be susceptible to it. The participant quoted in the passage below (T-C) was in their final year of high school, roughly 17 years of age, identified as male, and held a position of leadership with the school at the time of our discussion. Despite these characteristics and their laughter during the account, on review, it may be that social influence was a factor in this individual opening themselves up to anonymous comments (Huang & Li, 2016). In their own words: "All of mine were nice…[short pause]...except for one". Further conversation within the T-C group revealed that, while none of our participants who were younger referenced it, this group believed that their younger peers probably were using this anonymous platform, if not others.

T-C/ A. There's a lot of Year 7's using, like, they put it up as a harmless thing but they get...severely bullied on it. But then...they're proud to say it? From what I've noticed.

They'll fully put it on their Story and be like...it's like they're asking for it coz they're saying like "Go. Say something."

B. Yeah like, fishing for compliments and—

A. --Yeah, but they want the negativity...

C. The drama

A. ...to get sympathy [M & L agreeing]

D. Yeah it's kinda like..."do what you want, it doesn't bother me" or something...but...it does bother them.

E. Yeah but it always bothers people [several agreeing]. Especially younger people...

B. Yeah, no one's immune

In two separate focus groups from different locations and adjacent age groups (i.e., early vs. later high school years), participants shared strikingly similar examples of SNS being used as anonymous venues for critiquing school peers. A-M group participants spoke of their version of this, "[Their school's nickname] Secrets", which they all knew about but was believed to have been shut down by the site administrators the year prior. Among the group, pages like these were unanimously viewed poorly (e.g., "Yeah, [they] sucked!"). Despite the fact this page was expressly created for members of their school community, meaning that the users are likely to be known individuals, participants did not know the identities of the page creators or who the worst offenders for negative comments were. In examining the harmful effects of cyberbullying, Sticca and Perren (2013) suggest that being targeted by anonymous users can increase the level of harm, meaning that comments on the unofficial school page could be especially hurtful.

A-M/ No one goes into the depth to find out who you are, the things that you posted...like unless you're really, really offended.

Comparably, senior students (i.e., 17- and 18-year-olds) from one of the Tauranga-based focus groups (T-C) described their *The Most Likely To…* anonymous SNS page:

T-C/ M: Last year um, the Year 13s made an anonymous page for "The Most Likely To..."'s. And it was...like, you could write anything about someone else without them knowing...and that was really bad, there's was like...my sister was in that year and there was a heap that...everyone used it as a good opportunity to...[got quieter and then completely trailed off as that last sentence went on] // N: Yeah...pretty much roast someone.

Despite these possibilities for harmful impacts, it is crucial to view cyberbullying and acts of online aggression as a relatively small part of a much wider picture of internet use. Particularly when it comes to young people, online connectivity holds an important place in their day-to-day lives –a predominantly constructive place (e.g., Greenhow & Burton, 2011; Verduyn et al., 2017). Over and above their use of the core SNS, participants discussed a wide range of platforms to keep in touch with friends and family. I collated these SNS into an *instant message* category. Additionally, such sites were also used for establishing connections with like-minded others, including those with whom participants may otherwise never have crossed paths (e.g., Marciano et al., 2022). The potential benefits of connections like these should not be undervalued, particularly for anyone belonging to a group that is underrepresented in their face-to-face community (Mueller, 2019) or who may lack ready access to social resources (O'Dea & Campbell, 2011). Participants from most of my focus groups expressed this sentiment:

A-G/ And there's also a Japanese app that my friends...my friend uses, called Snow. It's similar to Snapchat...but it's not very popular coz it's Japanese-based...so yeah

H-C/ we [focus group participants] have a Group Chat and we use it for organising events and when we're going to meet up all that sort of thing.

A-M/ I use WhatsApp which is a very big thing in South Africa, where I came from.

T-C/ For me it's [Viber] a way like...like seeing and talking about someone's life on the other side of the world. I find that kinda thing really interesting.

Creating and sustaining meaningful connections through games that had a SNS component was also a key theme, which I categorised as *Games*. Communication functionality is commonplace in all manner of games, likely because most gamers are considered *socialisers* (Bartle, 2013). First proposed by Bartle (1996) and widely applied since, a two-by-two matrix (action vs interaction; and world- vs player-oriented) results in four player types. Socialisers, then, are characterised by valuing interaction and by being player-oriented, which integrates well with participants' views:

A-G/ A. Yeah! Yeah most games have it nowadays coz it like brings more people to the game

B. And you can invite people to play that game with you!

A. My friends are just huge Minecraft fans and they always....any spare time that they get, they usually just play Minecraft with each other...[laughing]...sounds kinda sad but yeah

The second kind of SNS that I collapsed into the games category were additional sites that participants used in conjunction with their games. For example, one participant from the T-C group referenced that they used Discord for "online gaming communication", which was "mostly used by people that you know, [to] play online games with". These adjacent platforms reportedly made it much easier to converse with their friends about their game-playing strategies. According to this participant, they and their friends had also made many new social connections purely via their gaming-related online activity. Furthermore, and beyond strictly virtual combinations of games and communication, some of our participants spoke of using gaming as a kind of collective social activity. That is, they would play multiplayer video games at an otherwise offline gathering as a way to bond (A-M).

A-M/ Well I mean, personally I don't play like shooter games, but I do play online games like FIFA and I was on eSports last year for FIFA '17. Doing a non-competitive one and like…we do it…coz I play for the [school's football team]. So we do things like FIFA nights and all that stuff, where we all get together and just, you know, play some video games and all that stuff because it's like… a good way to communicate with people.

F: It's sort of...it's more used for gaming. So online gaming communication, all that. So it's mostly used by, yeah, people that you know, play online games with other people. It's a lot easier to use, to be fair.

As was the case with instant messaging behaviour, focus group participants were adept at leveraging game-related SNS for their own social development. That the games, their content, and the challenges built into them can serve as easy entry points into conversation surely helps in building brand new connections (Bartlett, 2013). Indeed, Burke et al. (2011) found evidence that online environments appear to "level the playing field" between groups who are particularly out-going and social, to those who are less so. In multiplayer online games, of the types the A-M and T-C participants mentioned, this "levelling" seems linked to automatically having a shared goal and shared environment with the other players. Additionally, there is evidence from longitudinal studies that some multiplayer online games serve to improve the problem-solving skills of those who play them (Adachi & Willoughby, 2013); therefore, the act of collaborating with others to solve problems could further reinforce even newly established relationships.

T-C/ A. Um sometimes I play [multiplayer online games] with people that I do know, sometimes I don't

B. There are some pretty bad games…

A. See but it depends…it depends what games you play really… especially if people are like, there to play

Of the five categories of the lesser-used SNS, platforms that were foremost used for gathering content (i.e., *content*) were mentioned in all the discussions. The sites in this category ranged in terms of how content- vs. communication-focused they were viewed to be (e.g., Pinterest vs. Twitter, respectively), as well as how commonplace being anonymous compared to being identifiable was (e.g., Reddit vs. Twitter). That said, participants did still identify that all these SNS had the capacity for communication with others; therefore, they were compatible with the groups' established definition of SNS. Across focus groups, there was a consensus that participants used sites like these mostly passively to follow topics of interest. If they engaged in communication on these sites at all, as one participant put it, "you don't really communicate with people you know in the real world on it" (H-C). That communication *can* occur on all these sites does warrant some consideration from researchers of online social behaviour. Furthermore, many of the adverse impacts attributed to SNS broadly seem to be (although not conclusively; Valkenburg et al., 2021a) related to passive use rather than the active engagement more common in other types of SNS (Burke et al., 2011; Verduyn et al., 2017).

The final categorisation I used to summarise the SNS that participants mentioned was *video sharing*. There is a substantial overlap between this and the content- and games-related categories in that all sites had a primary function other than communication. YouTube was an SNS that was typical of the video-sharing category:

T-C/ A. I don't know if it counts as a social media thing...but YouTube's quite a big one, I think

B. Yeah...for girls especially with like, make up tutorials like...[H & L agreeing]

C. I think because you can use it for anything. So say if you're struggling in a class, you can get tutorialson there...but also you can...highlights of games, or specific...like specific people you like watching

B. It's good to watch when you're bored

The video-sharing category of SNS also had the capacity for more person-to-person types of communication in the comments sections. In each focus group, I used participant references to these video-and-comments-type SNS platforms segue toward a severity rating exercise which I included to guide the discussion on how participants interpreted different types of online communications. I presented the groups with 10 verbatim comments from YouTube, presented in Table 3, along with the average severity rating that focus group participants assigned the comment in aggregate. I selected the verbatim comments, from the corpus I had used in a previous study (Hodge & Sargisson, 2022), with the specific aim of encouraging participants to share their opinions on different types of online communication. Doing so was useful because discussing their perceptions of how others had behaved online led participants to share their own anecdotes more frequently than did only discussing the platforms. Furthermore, in response to asking their opinion on each comment, participants often shared how they personally would react if they saw a similar comment.

Table 3

Average Severity Ratings of Exemplar YouTube Comments

Rank Order	ID	Verbatim Comment	Average Rating
^ Most Negative	**5**	"[at Marwa Awil] Shut your stupid face. Why are you so stupis? Cant you see he lied all the way, the fucking deformed chimp" (A clip from a talent show, like America's Got Talent type thing)	1.00
	4	"He's asian so I'm not impressed."	1.57
	2b	"asains, they take things seriously"	1.71
	2a	"gangnam style is korean and he's chinese." ((Correcting a previous commenter))	2.46
	1	"man if you wanna learn to shuffle go to hip hop classes with the girls haahhaha"	2.62
	6	"The whole dance in general looks off from the beat? It kind of loks like someone slapped this song over the original."	3.00
	8	"She belongs in the movie step up 3" (A Dancing performance)	3.50
	2c	"we always have to mention someones race. Please, get a life, if ur gonna just add to the nonsense"	3.50
	7	"lets see you dance Stupid [sic], shes 9 years old give her a break"	3.67
	10	"hey kevin, man you're great for when you were just 11. don't get all bothered by people's negative comments. You just need to know that you are good and the rest just comes. So basically - screw people like Ohm. He most likely cant do even 10% as good as that."	4.33
	9	"um I think that u didn't do much dance but did more tricks then dance and u didn't flow with it u stped [sic] and started but u did really well :)"	4.50
Most Positive v	**3**	"This kid is going to grow up and be one of the best breakdancers. :D"	4.50

Comments 2a, 2b, and 2c were of particular interest due to being from a single comments thread (i.e., Comments 2a, 2b, and 2c). The first two comments from the thread (2a and 2b) tended to be rated as negative or moderately negative. In contrast, the response to these comments (2c) was rated more positively (e.g., A-M, A-G, & T-C). While several participants considered 2c to be a welcome rebuttal to the prior comments, they were critical of the tone, which highlights the complexity that can be present in online social behaviour (O'Reilly, 2019; Vermeulen et al., 2017; Wang et al., 2018).

A-M/ O: I had 3...because [of] the way he phrased it, he is going to get so much backlash!

A-G/ F: It's not like...yeah, he did it with good intention, but by doing that he just created so much more stuff

T-C/ P:...I try to put anything in the most positive way possible... I'd just try to put it in the most positive way possible.

Online comments are often difficult to interpret accurately due to incomplete contextual cues (e.g., Macháčková, 2012; Mittmann et al., 2022; Thornberg et al., 2018). Within the discussions, several participants referred to comments as being ambiguous and difficult to rate. One participant from the A-G group recognised that their own comments could be misinterpreted and said that they “make it [their intended meaning] clear with the laughing emoji…but if I’m being serious, like, I won’t”. Put another way, this participant compensated by adding a visual cue to help others read the intention behind their text-based comments. Researchers have also suggested that a certain amount of ambiguity online can be mitigated using emojis, particularly in cases where tone and non-verbal cues would ordinarily be used to convey meaning (Keane & Hammond, 2022; Losby, 2022). Conversely, participants from the P-A1 focus group, who were younger, had an opposing view. In response to a probing question regarding Comment 9, all three participants agreed that the emoji made little difference to how they interpreted the intention behind the comment.

<u>Researcher: What impact do you think the smiley face had on that [comment]?</u>

4/ P: Not much [laughing]. To me personally…a smiley faces [sp] or something [like] that doesn’t really make much difference

H: It’s like an illusion

G: Yeah

In the severity rating exercise, the difficulty of accurately assessing the intended meaning of text-based online social behaviour was perhaps best exemplified by responses to Comment 6: “The whole dance in general looks off from the beat? It kind of loks like someone slapped this song over the original”. As in Table 3, this comment received an average and neutral rating of 3, but the individual ratings covered a wide range of perceived severity, indicating that participants had mixed views about it. Comment 8 generated ample discussion about its author’s intentions due to the comparison point they used, the film *Step Up 3*. Within the same focus group, one person rated this as moderately positive while another rated it moderately negative, seemingly as a function of how they supposed the commenter viewed the film. “Step Up 3’s a good movie” and “Step up 3 is horrible” are respectively quotes from these raters. Comment 1 also garnered constructive conversation when it came to its severity, described as “in the middle [of the valence scale]” and “not to the point where anyone’s going to be really offended”. Comment 1 was seen as ambiguous and various participants noted that it “depends who said it”, “yeah, it kinda depends”, and “if it was your best friend … that could be a joke”. Also worth noting is that when one participant learnt the context was YouTube, rather than elsewhere online, this piece of context helped them decide on a severity rating: “Oh, [if it’s YouTube] then that’s negative!”

Participants were considerably more uniform in their interpretation of negative comments than they were in positive ones. Ratings were consistent when the comment made disparaging references to race, as in Comments 5 (M = 1, *Range* = 0) and 4 (M = 1.57, *Range* = 1, SD = 0.53, SEM = 0.20). The idea of intervening after someone posts a negative comment came up in six out of the seven focus groups. One P-A2 participant summed up the careful balance between intervening to “pile on” the negativity or to help defend someone by observing that “some of them are a negative dig at someone… and then someone else is backing them up. So it's...yeah, they're two different things”. All other references to intervening were conclusively in the direction of helping the person being targeted, which tends to be how most people respond – at least when asked (e.g., Leung et al., 2018).

In multiple focus groups, participants stated a variation of “You shouldn’t look at the comments” (A-M). Interestingly, the participant responsible for the quoted instance did go on to qualify this slightly (“Yeah I’ll read them… I just won’t interact with them”), to which another agreed, “Yeah good point, I’ll read them… I just won’t say anything”. This idea of looking but then trying to ignore content posted in a comments section is symptomatic of a broader idea; that information ignored is somehow neutralised. In actuality, the opposite is likely true. For example, when someone sees a hurtful post about themselves in an online public space, and if that post goes unchallenged by the other people presumed (by virtue of being online and public) to have seen it, there is evidence suggesting that any initial amount of harm is heightened (e.g., Dillon & Bushman, 2015). For the targets of the harmful post, not reading a comment is no guarantee that it won’t surface elsewhere; as a participant in the H-C group put it, “rumour spreads!” Moreover, because harmful behaviour like bullying is often related to social reinforcement (and the lack of social punishers) from others, bystanders ignoring an event may serve to maintain and condone such behaviour (Anderson et al., 2014; Antoniadou et al., 2019; Sarmiento et al., 2019).

Witnesses to harmful behaviour online (i.e., cyberbystanders) can make a meaningful impact both in supporting the target (Macháčková et al., 2012; Macháčková & Pfetsch, 2016) and in establishing a more constructive standard of behaviour in online spaces (Hogg & Reid, 2006; Tankard & Paluck, 2016; Wolter & Napper, 2023). Focus group participants shared several accounts of bystander behaviour they had been a part of or had witnessed. One of the first-hand accounts of bystander intervention was an A-M participant who helped to report what had begun as an in-person verbal assault, which later continued via SNS. The fact that the altercation moved online allowed for effective intervention because the participant “had already screenshotted what he said and we took it back to the school and all that". As part of the same conversation, the participant-bystander noted that the frequency of the insults toward their friend increased once social media became the venue. Another participant who had been privy to the online portion of this scenario asserted that "once it's known about, like 50 people can see it in two minutes", to which the group at large seemed to agree.

A second example of bystander intervention came from the H-C focus group. In this case, the harmful behaviour began in an online space and later became face-to-face when the participant confronted a bully on their friend's behalf:

> H-C/ F: One of my mates got into a really bad fight last year because...someone left like a comment on his thing, you know, like "I'll kill you, I'll come to your house, and I'll beat you" and all that kind of stuff. And... me and him went up to this guy and it was...he was like a, I don't know, 15 year old guy, and we confront him, like in real life. He started apologising but then the next day left the same comment.

The use of SNS for socialising purposes, rather than purely for entertainment, has been found to be most closely related to high levels of well-being (Wang et al., 2014). However, in the wider picture, there remains considerable disagreement between different researchers and their work on this issue. Researchers of one of the relatively few longitudinal studies in this area found that increased Facebook use lowered rates of subjective well-being (Kross et al., 2013). Participants from both the H-M and A-M focus groups seemed to describe a similar idea in that they spent more time than they would have liked on SNS because they were engaging passively and lost track of time.

> H-M/ So it's really bad for me coz I get home and I like…binge on my social media accounts [some others agreeing here]. I just have a really bad thing with…I just don't know how fast the time goes [lots of agreement from others here also]. I got off my phone and it was 1 in the morning this morning…this morning!

5/ S: Yeah like, I'll find I'll be on Facebook and I'll just be looking through stuff, or I'll be on Instagram looking at feeds and stuff…and um, I'll think I would've spent about ten minutes and…and hour has gone past!

To experimentally test the impact of time spent using SNS, Sagioglou and Greitemeyer (2014) and Tromholtet et al. (2015) assigned their respective participants to spend different amounts of time on SNS, including a group who took a week-long break. In both studies, the SNS-use conditions resulted in well-being measures that were statistically lower than the control condition. More in line with the experiences of two of our participants (i.e., in Groups P-A2 and A-G), Rae and Lonborg (2015) have suggested that the use of SNS like Facebook to maintain pre-existing friendships is strongly linked to high well-being, while less so with developing fresh connections online only.

P-A2/ Like I'm always…I don't…I'm not always on my phone but like…if someone texts me I reply straight away so…probably add up to a lot of hours in the end!

A-G/ Um, so I went on "the spirit of New Zealand" and half of them only have Facebook…so…and I've been wanting to get it for a while because so many people have asked me [to get Snapchat]…because that's the only social media they have…and coz I've got a whole bunch of friends in Singapore

Focus group participants suggested that the caregivers and school officials in their lives lean heavily on restrictive methods rather than active discussion and mediation alongside young people, as many researchers around the world have also documented (e.g., in the US by Bleakley et al., 2016; in Russia by Soldatova et al., 2020; in Brazil by Cabello-Hutt et al., 2017). The trend of traditional media sensationalism, which has accompanied advances like radio and television previously (Mueller, 2019), has similarly led to overly restrictive responses to the perceived threat of the online environment (Livingstone et al., 2017). In my research and that of others, young people express that the mere possibility of losing their internet access leads them to shy away from reporting incidents to the adults around them (Midgett et al., 2017; Sticca & Perren, 2013). Participants expressed exasperation over how online issues were responded to by their teachers (A-G), their school overall (H-C), and their parents (A-M).

A-G/ A. And [teachers] put a restriction on the website! So I can't check my marks, it sucks!

B. Yeah on the school website! They've got this massive logo…on the top of everything –

A. Yeah it's over everything! So if you click something…or if you go down here [scrolling down] and try…see! …they're just blocking sites for no reason.

H-C/ If you...so they block sites but if you have VPN you can access anything....because if one gets blocked you just go on the App Store and download another one. Then two weeks later, it's blocked and just rinse and repeat.

A-M/ I just kept going in and turning it on. Yeah and my mum would come in in the morning and be like "why is it on?" and I'd go "I don't know, I've been asleep". Turns it off...she walks out and I turn it back on. Then she came it like an hour later, she was like "why's it still on?" and I was like "maybe you forgot to turn it off properly". She walks out, I turn it back on...

An alternative interpretation, though, is that participant stories are indicative that they are anxious and hold little hope that their concerns would be addressed meaningfully:

P-A1/ Yeah but…the reason the teachers didn't do anything…coz they had…they had *something else better to do.*

Indeed, participants in focus group A-M detailed events that demonstrate something beyond failure to address a concern; rather, they indicate that these young people perceive adults to be naive of their young person's experience to the point that they worsen the situation.

> A-M/ A. …my parents didn't know how to [handle the cyberbullying incident]. So like...I tried...with the social media thing… I tried not to tell my parents but...every five seconds they'd be like "don't talk to them, don't communicate with them" and I kept explain[ing] to them, "I can't even do that coz of this, this, and this". They get so confused. They're like "just don't do this and make sure you do this" and I was like..."I've already done that...and I can't do that [other thing] anyway". And they just don't know how it works!
>
> B. Yeah and the blame also gets dropped on us *a lot*...but they don't understand that social media actually is...like it kind of naturally...once something happens, it just kind of erupts ... They don't understand that...all you have to do is send one text and...everything can come down!
>
> C. Yeah, like my parents...they like, if stuff happens with me on social media - which it has before - they never knew how to deal with it. They always went went like "just block them, don't talk to them". When...you can't do that.
>
> A. It's quite scary to think that like...we can deal with this situation where we're getting accused of something that could ruin our lives...that we could deal with it better than our parents could… coz it was actually so much more stressful with our parents every five minutes going "are you ok, make sure you do this". It's like...we are more capable of dealing with that situation than our parents are...scarily!

Increasingly, researchers have found that the use of restrictions for managing internet use results in a disproportionate reduction in the opportunities for young people relative to the possibility of experiencing harm (Cabello-Hutt, Cabello, & Claro, 2017). In other words, the likelihood of an adolescent encountering harm online does tend to reduce under restrictions;

however, the opportunities available to them do reduce to an even greater extent. One group of researchers put it definitively: “no combination of enabling and restrictive mediation both increases opportunities *and* reduces risks” (p. 97, Livingstone et al., 2017). Furthermore, restrictive measures also impinge upon young people becoming more digitally literate users on SNS - a factor that is a much better predictor of productive use of the internet (Rodríguez-de-Dios et al., 2018).

Summary

Participants in my focus groups discussed several proactive ways in which they were using technology to supplement their social live, across several different SNS platforms (RQ2.1). Although participants from every focus group were able to outline many benefits and opportunities that using SNS had afforded them, they also articulated struggling with self-regulating their own use of SNS (RQ2.2). In addition to general frustration about some types of online social behaviour (e.g., trolling, rumours spreading, etc.), participants also shared stories of serious harmful conduct that they had been involved in and they had felt poorly supported in these situations (RQ2.3).

Overall, the results from my focus group research supports the assumption that cyberbystander intervention can help minimise the effects of harmful behaviour and, therefore, warrants further research in pursuit of a more inclusive online environment. The severity rating exercise and resultant discussion highlighted that, in many cases, participants found it difficult to determine the intentions behind behaviour which often led to them to interpret online comments as ambiguous. Additionally, participants spoke about how anonymity tended to lead to behaviour that was more severe, as well as making it more difficult to seek the help of adults. Therefore, how ambiguity and severity (e.g., that may be heightened with anonymity) impact upon cyberbystander intervention emerged from my focus group study as a promising direction for further research.

Chapter 3: Connections

Outline

In this chapter, I outline the problem of aversive behaviour on social networking sites (SNS) and the relevant research on the topic to date. Further, I focus on the type and frequency of behaviour intended to support others online, such as cyberbystander intervention which was raised by several focus group participants (Chapter 2). Following that, I present an alternative theoretical and methodological approach to complement the extant SNS research.

Part 1: The Problem

Social Networking Sites

The increasing usage of social networking sites (SNS) has led to concerns regarding its impact on mental health and social connectedness (e.g., Cunningham et al., 2021; Moreno & Whitehill, 2014; Twenge et al., 2018). Many research studies have examined the relationship between SNS usage and an individual's mental health (e.g., Coyne et al., 2020; Odgers & Jensen, 2020; Schønning et al., 2020). Some researchers have argued that excessive use of SNS can negatively affect social connectedness, particularly among adolescent populations (e.g., Easton et al., 2018; Mishna et al., 2017; Sagioglou & Greitemeyer, 2014). The notion that SNS usage replaces face-to-face interaction with friends has contributed to this negative perception (Twenge et al., 2018). However, research from others has challenged the view that socialising online necessarily displaces face-to-face connections (Boer et al., 2021; Rosenfeld & Thomas, 2012). The debate regarding the overall positive or negative impact of SNS usage on social connectedness has been present since the inception of this field of research (Kiesler et al., 1984). Moreover, this debate regarding SNS being net good or not remains nuanced today (Orben & Przybylski, 2019).

In one of the first longitudinal studies on internet use for social purposes (Kraut et al., 1998), researchers found some evidence of isolation among high-frequency users. However, despite subsequent efforts by researchers, including Kraut et al. (2002), to revise this initial conclusion, the negative perception of SNS has persisted. A follow-up study by Kraut et al. (2002) found that the relationship between SNS use and isolation was substantially weaker than in their earlier work (e.g., Kraut et al., 1998). In fact, SNS use appeared to increase individuals' offline community involvement (Kraut et al., 2002). Other longitudinal work suggests that the perceived negative link between SNS use and well-being may be unwarranted (e.g., Heffer et al., 2019). It is worth noting that the original findings (e.g., Kraut et al., 1998) were based on a population that was relatively early adopters of the internet and SNS. Thus, the early indication of isolation may have been due to the limited pool of online conversation partners rather than any inherent isolation associated with the form of communication (Williams, 2006).

The negative perception of SNS use is an issue because SNS is a critical component of many individuals' everyday communicative behaviour. For example, SNS can facilitate friendships over long distances (O'Reilly, 2020), strengthen existing connections (Burke et al., 2011; Cheng et al., 2014), and be a venue for young people's identity formation (Peña-Fernández et al., 2023). SNS can also provide opportunities for interaction with more diverse groups of people (O'Dea & Campbell, 2011), partly because it is uncommon for rare individuals (i.e., rare in some dimension) to be geographically close to others who are similar. Consider someone who has a rare medical diagnosis. By definition, it is unlikely that this individual would meet others with the same diagnosis without SNS and the ability to seek out groups based on wide-ranging mutual interests (Marciano et al., 2022).

Furthermore, SNS may be beneficial for those at risk of being marginalised (Mueller, 2019). That is, the malleable structure of online spaces may make it easier to explore less represented identities. For example, a literature review showed how Indigenous Australians used SNS to strengthen their identity, gain confidence across multiple arenas, and further engage in research spaces where they were otherwise underrepresented (Rice et al., 2016). Similarly, researchers have found that using SNS can benefit areas like one's professional life (Utz & Breuer, 2016) and, more generally, enhance overall social capital (i.e., gaining social resources; Kraut et al., 2002; Williams, 2006).

The relationship between SNS use and a corresponding decrease in mental well-being is inconclusive (e.g., Jensen et al., 2019; Orben & Przybylski, 2019; Valkenburg, 2021). In fact, some evidence suggests no relationship between SNS use and long-term mental health outcomes, nor between SNS use and mental health on any given day (Jensen et al., 2019). Specifically, the frequency of technology use did not predict later mental health outcomes or vary as a function of daily mental health status (Jensen et al., 2019).

It is possible that measuring the frequency of SNS use is not enough to predict mental health. Some researchers suggest that differentiating between active and passive SNS interactions could provide a more nuanced understanding of the relationship between SNS and well-being (e.g., Quiroz & Mickelson, 2021; Verduyn et al., 2017). In this view, active users are those who engage with others online in a way that improves their well-being and strengthens their connections. Conversely, passive users are characterised as those who spend a similar amount of time online but mainly observe the interactions of others rather than cultivating their own. This theory of active and passive SNS use suggests that passive users are prone to comparing themselves unfavourably to the idealised lives they see on SNS; thus, they are more likely to experience poor well-being (Quiroz & Mickelson, 2021).

Nonetheless, research on the extent to which dichotomising SNS use helps to explain young people's well-being is still unclear (Orben & Przybylski, 2019; Valkenburg, 2021; Valkenburg et al., 2021b). Another related theory is that some people may have a maladaptive reliance on social technologies, even if they spend less time online overall than others (Boer et al., 2021). Put another way, the duration of time that some individuals spend online may not be the source of the issue.

Aversive Online Behaviour

Discussions and research surrounding social behaviour on SNS often encompass bullying and other similarly negative behaviours (Lysenstøen et al., 2021). Bullying is often characterised by a power imbalance between the target and the perpetrator, the intent to cause harm, and repetition of the behaviour (Olweus, 1997). That said, even in face-to-face bullying, it can be challenging to determine the intentionality of negative behaviour, as the absence of intention does not necessarily prevent harm from being inflicted. As an example, misinterpreting what is intended to be a joke can cause harm, and, equally, a lack of self-awareness on the part of a bully could exclude a behaviour from categorisation as “bullying” (Thomas et al., 2017). In much of the extant cyberbullying literature, the same definition used for face-to-face bullying is employed, but for the difference that the behaviour occurs in an online space. Therefore, the definition of “bullying” and “cyberbullying” share many of the same issues (Olweus & Limber, 2018).

In this thesis, I use the term “aversive behaviour” to include all behaviours that cause perceived harm to others, including but not limited to cyberbullying. The broader term of aversive behaviour is likely more functional, as young people do not typically include intentionality in their understanding of bullying (Byrne et al., 2016; Vaillancourt et al., 2008). Moreover, decision-making about addressing aversive behaviour does not gain anything by categorising the behaviour as “bullying” (Ross & Horner, 2009).

Definition issues notwithstanding, there remains satisfactory empirical evidence that cyberbullying is a problem; thus, so too is the broader category of aversive online behaviour. In a meta-analysis, Kowalski et al. (2014) estimated the prevalence of being a victim of cyberbullying to be within the range of 10-40%. This broad estimate of cyberbullying appears still to be representative of the extent of the problem. More recent meta-analytic studies found a prevalence range of approximately 14-58% for those attending school (Zhu et al., 2021) and between 8-28% for young adults (Kowalski et al., 2019). Suffice it to say that aversive forms of online behaviour are and continue to be a social problem that warrants attention.

The phenomenon of cyberbullying and other internet-mediated poor behaviour is sometimes characterised as the result of specific technological platforms (e.g., SNS) or advancements (e.g., smartphones or video messaging; Catone et al., 2020; Sung-min & Byong-jin, 2022). However, the consistent underlying factor is human behaviour. It is, therefore, vital to explore the functions of online social behaviour and the environmental factors that shape or maintain it.

Part 2: Supportive Behaviour Online

According to behaviourists, behaviour is influenced by the environment in which it occurs (Baum, 2017; Skinner, 1953), which includes behaviour that is reinforced or punished by social outcomes (e.g., Chen et al., 2014; Guerin, 1994; Qin et al., 2023; Safin et al., 2015). By definition, SNS are online social environments. That is, SNS is considered to be any space (i.e., "site") in which groups of people (i.e., "social networks") can communicate with one another. As with other socially mediated environments, there is the potential for harmful behaviour, such as that outlined in Part 1 (Lysentøen et al., 2021). Within the findings from my focus groups (Chapter 2), participants acknowledged that their use of SNS was complicated and, at times, negatively impacted their well-being. In other words, participants recognised that they and their peers encounter psychologically aversive situations on SNS. In response, they were interested in how online support can manifest. Others' participants echo the insights from my focus group participants (e.g., Luo & Bussey, 2022; Metcalfe & Llewellyn, 2019; Mishna et al., 2018), as do many researchers (e.g., Lysentøen et al., 2021; Macháčková et al., 2012). Furthermore, when Lysentøen et al. (2021) sought to review empirical research on social media use and prosocial behaviour, they found that such literature was scarce.

Bystander Intervention

One way to support others is to intervene when a peer is the target of bullying, mocking, or any comparably aversive behaviour. Intervening disrupts the aversive behaviour (Olweus, 1997; Salmilvalli, 1999). Intervening may also stop aversive behaviour entirely (Barnett et al., 2019) or, if not, mediate harmful effects for those targeted (e.g., Banyard, 2008; Bastiaensens et al., 2014; Macháčková & Pfetsch, 2016). As an example, one of my focus-group participants described an instance where they intervened for a friend. The participant (and a friend) verbally confronted someone who had been perpetrating aversive behaviour on an SNS space usually used for social organisation between schoolmates. From the participant's perspective, the in-person confrontation caused poor online behaviour to stop (Chapter 2). Similarly, researchers have found that the visible presence of participative bystanders (who do not necessarily intervene) can reduce the likelihood of conflict arising and decrease the chance of conflict escalation if it does occur (Lassiter et al., 2021). Moreover, the actions of those not directly implicated are likely more malleable than those of targets and instigators of aversive behaviour (Cassidy et al., 2013). Therefore, whether onlookers intervene in aversive behaviour is a critical and complementary component to extant work regarding minimising harmful social behaviour such as bullying (Allison & Bussey, 2016).

Bystander intervention is the term most often used by researchers interested in the prevalence and motivations for behaviour like that in the example above. A "bystander" is typically understood to be someone who witnesses an event (Latané & Darley, 1968). Further, the term bystander suggests other behavioural options beyond standing by as a witness (Paull et al., 2012). The bystander who intervenes is referred to as an active bystander, or sometimes a defending bystander (prosocial, e.g., helping a target) or a bystander supporting the aversive behaviour (antisocial, e.g., supporting a bully; Ng et al., 2022). The bystander who does not intervene is usually called a passive bystander (e.g., Bastiaensens et al., 2014; Ng et al., 2022). Some researchers have contended that there are two subsets of passive bystanders (Paull et al., 2012). For my purpose, it was critically relevant that both types of passive bystander behaviour are indiscriminate from the point of view of someone experiencing harm. One type of passive bystander considers that someone should intervene, although they do not. The other type of passive bystander does not consider intervening at all. Neither type intervenes; thus, their behaviour is identical to a target and any bystanders. For that reason, I consider all non-intervening bystanders to be passive bystanders for this thesis. Although there are several sub-categories of bystander intervention, Leung et al. (2018) found that intervening to defend a target of aversive behaviour was considered the normative response by the majority of their participants, including those who had perpetrated bullying.

The term bystander is sometimes used to refer to a person on the whole rather than to call attention to a set of behaviours exhibited in certain situations. It is important to note that a person can act as a passive bystander in one situation and then behave as an actively defending bystander in another (Safin, 2018). Under the right circumstances, most individuals can exhibit prosocial behaviour as active bystanders (e.g., Baldassarri & Grossman, 2013). Put another way, being a passive bystander is not a fixed trait but is highly dependent on each situation's context. Nonetheless, in this thesis, I will keep my terminology consistent with the current work on this topic (Fischer et al., 2011). Specifically, I will use the "bystander" to refer to helping defend a target from otherwise aversive behaviour. In cases where there is an opportunity to intervene, but someone does not, I will specify that person as a "passive bystander" in that circumstance.

Measurement of Bystander Intervention

Researchers have approached the measurement of bystander intervention in several ways. Most common to date are self-report methods, such as asking whether someone has previously intervened in an aversive scenario. Questions like these that are direct, that strongly indicate a "correct" answer, and are used as standalone measures (rather than a scale of several items) are especially prone to the bias of social desirability (Ryff, 1989). Social desirability is when participants behave in the way they perceive as optimal rather than as they ordinarily would when unobserved (Crowne & Marlowe, 1960). Prosocially intervening to support others is likely desirable behaviour (Lanz et al., 2021; Roussel, 2017). When researchers ask participants whether they have intervened to help another person, some individuals may respond that they had intervened in the past, even when they had not. Therefore, I interpret single-item measurements of bystander intervention cautiously, as they may overestimate the actual prevalence of such behaviour.

Another approach is to observe whether bystanders intervene. In early bystander effect studies, semi-naturalistic scenarios were commonly set up on university campuses with the help of confederates (i.e., someone playing a role). Darley and Latané (1968) carried out several such studies, such as when they facilitated a conversation over an intercom between one student participant and several confederates posing as students. Part-way through the group discussion, a confederate feigned an epileptic seizure. The critical study measure is the duration between the onset of the seizure and when the participant notified the research staff of the emergency. Darley and Latané (1968) found that the duration between onset and the participants seeking help was longer when the number of other people ostensibly participating (i.e., the number of bystanders) was larger. Latané and Dabbs (1975) found that the observed participant was less likely to help as the number of other bystanders increased. That is, their helping behaviour seemed inhibited in the presence of others.

Researchers have documented that situational ambiguity also reduces the likelihood of bystanders intervening (e.g., Latané & Rodin, 1969). Situational ambiguity is when some aspect of a situation is unclear in such a way that it proves challenging to interpret what is happening and what one should do. Situational ambiguity can be a product of several factors. For example, there may be ambiguity regarding who is involved and how they are related to one another due to anonymity (e.g., Solomon et al., 1982). Furthermore, a situation could be ambiguous if it is unclear to observers whether a scene is being acted out in jest. Consider a case where an argument begins between audience members during an audience-interactive theatre production; it would be difficult to assess whether such a case was severe and required help or was intended as part of the performance (Phares & Wilson, 1972).

While observing how people respond in semi-naturalistic situations is more likely to reflect an individual's actual behaviour than in self-reports, it is essential to note that research like this can be constrained (Lambertz-Berndt, & Allen, 2017). Foremost because to create realistic and interactive situations in which to observe social behaviour, other people (or the suggestion of them) would generally be present. For example, in most of the studies mentioned above (i.e., Darley & Latané, 1968; Latané & Dabbs, 1975; Latané & Rodin, 1969), confederates were instrumental in the research design. By definition, confederates play a specified role in the study, while participants are led to believe the confederate is a fellow participant (Lambertz-Berndt, & Allen, 2017). Thus, using a confederate means that researchers deceive participants for at least some portion of their study. Although it is ethically permissible to deceive participants in some circumstances, research designs that do not require obfuscation are strongly preferred (Stark, 2010). In other words, if an alternate design is likely to produce comparable results without using deception, the alternate design should be utilised.

Hypothetical scenarios have become crucial methods for researchers interested in topics where their work would otherwise have been constrained (Bickel & Marsch, 2001). Hypothetical scenarios have been used to investigate a wide range of behaviour. For example, Foreman et al. (2019) used vignettes to explore participants' views and behaviour regarding texting and driving. Others have combined hypothetical scenarios to examine differences in concern versus action on climate change (Rinderhagen & Sargisson, 2021), the influence that peers have on risky decision-making (Andrews, 2022), health-related behaviour (Friedel et al., 2014; Johnson & Bruner, 2012; Lawyer & Schoepflin, 2013), and many others (see Rung & Madden, 2018 for further examples). Moreover, Mazur (1987) suggests that in many cases, data collected using hypothetical scenarios is comparable in direction and trend (if not also in magnitude) to actual behaviour.

Research designs where hypothetical scenarios are presented to all participants are also useful for estimating the prevalence of passive bystanders, which can otherwise be challenging to determine. To illustrate, consider the work of McDonald et al. (2016) in their review of bystander behaviour in the workplace. Employees of the workplace were readily contactable, and therefore McDonald et al. (2016) estimated that passive bystander behaviour was a factor in at least 66% of cases of reported sexual harassment. In the context of SNS, though, the number of bystanders present and able to observe aversive behaviour in an online space is often unclear (Tian, 2016). As employee personnel files provided McDonald et al. (2016) with an upper limit for how large the number of bystanders could be, so too does the sample size of a study where all participants read and respond to hypothetical vignettes. The number of other people at an event is important because active participants are evident and easily counted. Using the total number of bystanders, those who do not actively participate (i.e., passive bystanders) can be calculated by subtracting the active participants from the total number of people. The total number of bystanders is usually not known for naturalistic events that are simply observed (rather than manipulated), meaning that measuring the number of passive bystanders is beyond a researcher's reach. In contrast, the total number of bystanders in hypothetical scenario research is known as it is the same as the study sample size.

The Bystander Intervention Model

While there has been much psychological research on how the bystander effect (Latane & Darley, 1970) functions in various offline scenarios (Fischer et al., 2011), more needs to be done to assess whether the effect similarly translates into online altercations. I explored online bystander behaviour through the relatively novel (to the research area) behavioural discounting method. In particular, I compared my discounting results to the five-step bystander intervention model presented in Figure 1 (BIM; Latané & Darley, 1970).

Figure 1

The Five-Step Bystander Intervention Model (BIM) as Proposed by Latane & Darley (1970)

Step 1. Notice the Event

→ Step 2. Interpret the Situation as a Problem

→ Step 3. Assume Personal Responsibility

→ Step 4. Know How to Help

→ Step 5. Help

Researchers have used the BIM to great effect in behavioural change regarding several forms of relational violence (e.g., Burn, 2009; McMahon & Dick, 2011) and as a protective factor against suicidal ideation (e.g., Worsteling & Keating, 2022). Regarding cyberbullying specifically, Ferreira et al. (2020) applied the BIM to the behaviour of bystanders. Those bystanders with higher self-efficacy were less likely to respond with further aggression and, instead, executed a prosocial intervention strategy (Ferreira et al., 2020). Researchers have suggested that incident severity and gender identity are related to the likelihood of bystander intervention online (Bastiaensens et al., 2014). Others posit that a larger audience reduces the chance of intervention (Macháčková et al., 2016). Moreover, Allison and Bussey (2016) provide a comprehensive overview of other contextual factors implicated in bystander decision-making, such as peer influence and the technological features of the online space in which the situation occurs. Help from even one of many possible bystanders substantially ameliorates the harm experienced by targets (e.g., Van Heugten, 2011); therefore, any variables that could increase the likelihood of active intervention warrant research attention.

Cyberbystanders

As interest in internet-mediated communication and online bystander behaviour grew so did the vocabulary used by researchers examining it. Specifically, online bystanders came to be referred to as “cyberbystanders” (Barlińska et al., 2015; Palasinski, 2012).

Cyberbystanders are defined the same way as usual bystanders, except that cyberbystander behaviour occurs online (e.g., Dillon & Bushman, 2015; Macháčková et al., 2015; Price et al., 2013). Therefore, I use cyberbystanders as a general term henceforth, and when someone does not intervene, I specify that they are behaving as passive cyberbystanders.

Research about cyberbystander behaviour is substantially newer than offline bystander behaviour. That said, there are some instructive studies from the mid-2000s. For example, Blair et al. (2005) examined how bystanders responded to a plea for help received in the form of an email. They found fewer responses to the email when many other recipients were copied in; thus, they found support for the bystander effect. The relationship between the number of bystanders who witnessed the email (i.e., the other recipients) and a participant's willingness to offer help followed a hyperbolic function. The difference in responses when there were few bystanders (e.g., one person compared to two people) was much greater than when comparisons between groups of 15 and 50 others were made (Blair et al., 2005). For the study group who read about someone else offering help in a similar situation (i.e., who were prosocially primed) prior to participating, the number of other bystanders had little impact on responses. In contrast, the responses of those not primed toward helping behaviour tended to show a hyperbolic decrease in their inclination to intervene as a function of the increasing audience size. In sum, Blair et al. (2005) provided an early example of how audience size may influence cyberbystander behaviour and an indication that a hyperbolic function could be a useful explanatory model.

In addition, the results from Blair et al. (2005) raise the critical question of how social norms impact cyberbystander behaviour. In the context of that study (Blair et al., 2005), the behaviour was partly changed by the participant learning about the prosocial act of another person immediately before being presented with a scenario in which they could similarly offer to help. Put another way, helping behaviour had recently been modelled for participants, and likely strengthened the perception of such a response (i.e., helping) as normative (Allison & Bussey, 2017; Leung et al., 2018). Subsequently, the salience of that behavioural norm (i.e., helping others) means that individuals will likely behave according to that perceived norm (Hogg & Reid, 2006). Behaviourally, then, social norms act as antecedent interventions. In a complementary study, Parris et al. (2020) examined the influence of a lack of prosocial intervention. The prevalence of aversive behaviour was greatest when increased numbers of bystanders remained passive. Moreover, aversive behaviour often continued longer than in schools with more active interveners (Parris et al., 2020). In sum, if intervening to stop aversive behaviour is normative, then aversive behaviour decreases (e.g., Bastiaensens et al., 2016; Blair et al., 2005). If the norm is that people do not intervene to help targets (i.e., intervening is not normative), then aversive behaviour seems to proliferate in that environment (e.g., Parris et al., 2020).

It is important to note that the social norms themselves are not the researcher's primary focus. Instead, the focus is on people's perception of social norms (Tankard & Paluck, 2016). The average person does not know the true rate at which cyberbystanders intervene to defend targets of aversive behaviour; thus, the actual rate of behaviour is not a factor in their decision-making. On the other hand, an individual's perception of how frequently other cyberbystanders intervene can act as a discriminative stimulus for their intervening behaviour. If people believe that most other people would intervene, they are likelier to do so (Bastiaensens et al., 2016; Deutsch & Gerard, 1955). What behaviour is perceived to be normative by an individual is informed by several factors, only one of which is the rate at which they have observed the behaviour of others (Allison & Bussey, 2017). Therefore, changing people's perceptions of normative behaviour has been posited as a way to effect social change (Tankard & Paluck, 2016). Moreover, strong social norms perceived relatively homogenously reduce situational ambiguity, as there are clear expectations about how one should behave (Mu et al., 2015).

The inverse, when there is no evident social norm, results in ambiguity (Alisson & Bussey, 2016). The ambiguity of whether a scenario requires bystanders to intervene is a core concept of the bystander effect theory. According to bystander effect theory, the more ambiguous a scene is, the less likely a bystander will intervene to assist (Latané & Darley, 1970). As in offline circumstances (e.g., Darley & Latané, 1968; Hawkins et al., 2001; Solomon et al., 1982), various factors in online spaces can lead to a situation that is difficult to interpret and, thus, ambiguous (Bastiaensens et al., 2014; Ferreira et al., 2020).

The level of anonymity that online spaces can provide is one factor that impacts the ambiguity of a situation. Small differences in the perceived anonymity of others can decrease the chance that someone will behave prosocially (Chen et al., 2022). For example, anonymity can decrease cooperation in social dilemmas, such as the Prisoner's Dilemma Game, where the identity of the players who made cooperative choices and those who made selfish choices is obscured from participants making a similar choice (i.e., unable to be identified, Komorita et al., 1992). Similarly, prosocial behaviour tends to decrease even when participants are asked to make considerably more arbitrary choices, such as which avatar to pass a virtual ball to in a simple ball-toss game (e.g., Hartgerink et al., 2015; Williams & Jarvis, 2006). However, blaming the advent of aversive online behaviour solely on anonymity is too simplistic, as each user’s history of reinforcement in anonymous spaces will vary (Dyer & Abidin, 2022). It is also possible for aversive behaviour to occur in non-anonymous spaces (e.g., Facebook; Rosenthal et al., 2016), and interactions on anonymous platforms do not necessarily result in aversive behaviour (Matias, 2017). The underlying mechanism that links anonymity and decreased prosocial behaviour is not fully understood, although self-preservation may play a role in such games and in cyberbystander decision-making (Alisson & Bussey, 2016; Dillon & Bushman, 2015).

Some may argue that intervening in online aversive behaviour puts oneself at risk - the antithesis of self-preservation (e.g., Abbate et al., 2022). By intervening, a cyberbystander could become the new target of aversive behaviour; thus, the cyberbystander is met with a form of positive punisher (i.e., retaliation). In a school setting, Rivers et al. (2009) hypothesised that students assumed they would become an aggressor's next target if they interrupted their bullying behaviour. To ameliorate the perceived threat by intervening, a bystander could avoid such situations. Avoidance constitutes a negative reinforcer due to removing themselves from a chance of harm (Dymond et al., 2014; Schlund et al., 2021). A substantial body of research suggests that in the face of threatening behaviour, a sizeable proportion of people exhibit avoidance behaviours (Dymond & Roche, 2009; Schlund et al., 2021; Sidman, 1955). Furthermore, high situational ambiguity can raise the threat level of a situation, even when comparatively little threat is present (Alisson & Bussey, 2016).

Although several situational factors influence the likelihood of cyberbystander intervention, researchers of much of the extant literature examine personality traits. Researchers in social psychology and personality psychology (in addition to others) have contributed substantially to understanding bystander and cyberbystander behaviour (e.g., Fischer et al., 2011; Lei et al., 2020). At first look, some social psychological concepts are challenging to explain within the framework of behaviourism. In general terms, I reconcile concepts that appear to require introspection (e.g., self-esteem) with a behavioural account by viewing these “traits” as instances of behaviour that have historically been reinforced or punished by others (Baum, 2017). That is, public behaviour and one’s history of having that behaviour observed and reinforced (or punished) is sufficient to explain constructs that involve introspection (Baum, 2017; Bem, 1967; 1972). With this definition of personality traits as a reference, several constructs aid in understanding cyberbystander behaviour.

Firstly, the construct of self-esteem. Self-esteem is described as one's internal self-regard; therefore, high self-esteem can usually be recognised as an individual expressing confidence (Coopersmith, 1967; Rosenberg, 1965; Rosenberg et al., 1995). Those with high self-esteem are more likely to have received praise from others more consistently than those with low self-esteem (Brockner & Hulton, 1978). Such praise may or may not be causally connected to a more objective measure of the individual's performance. Consider someone in business who had inherited their wealth and success rather than having established their status themselves. The employees of this businessperson are likely to give them high praise, but this may be related to the power difference between the two rather than strictly being related to whether they are good at their job.

There is a sizable amount of research concerning bullying and self-esteem (Lei et al., 2020). Targets of bullying are typically lower in self-esteem (e.g., Burger & Bachmann, 2021), have less social capital available to them (e.g., fewer friends to support them; Cramer, 1990), and are more at risk of becoming isolated from others (Kearney & Silverman, 1995). Researchers from related fields have similarly found that being the target of socially aversive behaviour, such as being ostracised, is related to lower levels of self-esteem (Williams, 2006).

The psychological construct of self-efficacy has similarities to that of self-esteem. Self-efficacy is the belief that difficult situations are challenges that can be overcome rather than threats to a person's confidence that should be avoided whenever possible (Bandura, 1982). Self-esteem and self-efficacy are related because the first is about whether someone has a positive history of reinforcement such that they have confidence in their abilities (as confirmed by the praise of others). The second is about whether someone has the ability and skills to approach problems proactively rather than losing confidence (i.e., decreasing self-esteem) when confronted. Self-efficacy is more likely than self-esteem to be examined within particular domains, such as academic (e.g., forming productive study habits; Gore et al., 2005), emotional (e.g., regulating self and others' emotions; Kirk et al., 2008), and social (Van Cleemput et al., 2014). In terms of the latter, several studies have shown that adolescents who typically help targets of bullying possess higher levels of social self-efficacy (e.g., Caravita et al., 2009; Pöyhönen et al., 2012; Van Cleemput et al., 2014).

Behaviourally, both self-esteem and self-efficacy can be considered products of an individual's history of reinforcement (Biglan & Lichtenstein, 1984; Freeman & Morris, 1999). Using a contextual social analysis approach (e.g., Guerin, 2016), high levels of self-efficacy are established through reinforcement for completing tasks. On the other hand, high levels of self-esteem seem to be established through iterative reinforcement of one's competencies and experiences (Guerin, 1994, 2016). Put another way, when an individual's successes are appropriately recognised and celebrated, others are signalling that the individual is held in high esteem. In turn, these conditions act as a discriminative stimulus such that if the individual recognises (i.e., discriminates) that they are held in esteem by their social peers, they then self-identify as having high self-esteem (Tajfel, 1982) and are more likely to behave in ways typically categorised as "high self-esteem" (Roche et al., 2002).

High self-esteem and self-efficacy are implicated in bystander intervention at several points, as conceptualised within the framework of the Bystander Intervention Model (Latane & Darley, 1970). Firstly, in Step 2, bystanders' judgement of whether the situation requires intervention can be influenced by their level of self-esteem (Witte, 1992). Secondly, self-esteem relates to Step 3 of the BIM (Latane & Darley, 1970) in which participants take responsibility for intervening rather than expecting someone else to do so, which requires the individual to feel capable (Paulhus & Martin, 1987). Thirdly, self-esteem is part of Step 4 that necessitates a bystander knowing how to intervene and, again, feeling capable (Paulhus &

Martin, 1987). In other words, to intervene one must feel a high enough sense of self-esteem and, relatedly, a great enough belief in their self-efficacy regarding intervening. Finally, in Step 5 of the BIM where the action is undertaken, self-esteem is an important factor because carrying out one's intentions successfully is linked to higher levels of self-esteem (McCarty et al., 2016).

Related to self-esteem, previous researchers have proposed a link between time spent online and a decrease in general well-being (Twenge, 2017; Twenge et al., 2018; Vannucci et al., 2019). Although, others contest that this premise is incomplete and requires distinctions between how the time online is utilised to be useful (Verduyn et al., 2017). Moreover, many researchers have found that accurately measuring time spent on the internet is troublesome (e.g., Wilcockson et al., 2018; Winskel et al., 2019). For example, Scharkow (2016) tested the agreement between participant self-reports of online time and the log files from their devices over the same period, finding only a moderate correlation. Contrary to researcher expectations, the direction of the correlation was such that participants tended to overestimate their time spent on various online activities. Scharkow (2016) additionally suggests that self-reports of time spent on social media platforms are comparatively more accurate than those of general internet use. Other researchers were more trusting of self-reports of screen time, having found that participant estimates were accurate to within approximately 30 minutes of the true value (Schuster et al., 2022). If such measures are valid, individuals' time spent online may be related to their cyberbystander behaviour in some of the same ways as self-esteem, such as time online inhibiting judgements about which scenarios require intervention (e.g., Step 2 of the BIM). At a minimum, time spent online and on SNS is a prerequisite for being a cyberbystander. That is, to witness online (i.e., cyber) aversive behaviour, one must first spend time online. It also follows that spending more extended periods online increases the chance and the frequency of being a cyberbystander to aversive behaviour (Park et al., 2014).

In addition to self-esteem and time online, an individual's socioeconomic status might be related to whether cyberbystanders intervene in aversive behaviour. More precisely, predicting whether an individual will intervene could be related to their perception of their socioeconomic rank within society, which can differ in important ways from objective socioeconomic measurements (Adler et al., 2000; Kraus et al., 2009). One view on how perceived socioeconomic status interacts with social behaviour is that lower-status individuals are more likely to empathise with others in hardship (Dubois et al., 2015; Stellar et al., 2012) and to behave more charitably toward others, including being more likely to offer others help (Piff et al., 2010). However, some researchers find the opposite; those who consider themselves comparatively worse off than others are more concerned with self-preservation and do not have the capacity to behave prosocially in the interest of others (e.g., Korndörfer et al., 2015; Stamos et al., 2020). As a type of prosocial behaviour, the form of relationship between a bystander who intervenes and that bystander's perceived socioeconomic status could help inform the debate regarding social status and prosociality, in addition to contributing to explanations of cyberbystander behaviour.

Behavioural Analysis

To summarise the ideas in Part 2 behaviourally, cyberbystander intervention can act as a negative punisher toward an aggressor (by removing peer attention) and is a critical component in minimising harm experienced online. Behaviours such as bullying tend to be maintained by peer attention and the favourable appraisal of those peers (Ross & Horner, 2009; Salmivalli, 2002). Intervening to support the target of bullying removes the attention from the bully such that the bullying is no longer reinforced; thus, the behaviour is negatively punished and should reduce. In school-wide interventions where students were trained to withhold social reinforcement of bullying (e.g., attention), follow-up observations indicated a 72% decrease in bullying-related behaviour (Ross & Horner, 2009). In other words, removing social attention for bullying behaviour reduced its prevalence. Additionally, modelling effective bystander intervention deters poor behaviour by changing what is considered normative (Allison & Bussey, 2017; Mu et al., 2015) and clarifying to others what an appropriate response to aversive behaviour would be (i.e., reducing the ambiguity around how to respond; Bastiaensens et al., 2016; Dillon & Bushman, 2015).

Part 3: An Alternative Methodological Approach

Exponential Discounting

Discounting was first applied in economics (Fisher, 1930) where researchers are concerned with how people make decisions about money and resources. Researchers in the field of economics often use an exponential function to describe such decision-making (Loewenstein & Thaler, 1989). Within this framework, the exponential formula can be represented as:

$$Discounted\ Value = Initial\ Value \times (1 + r)^{-n} \tag{1}$$

Where *r* is the discount rate, or, the estimate of the opportunity cost of using money in the present rather than investing in potential future benefits. *n* represents the units of time over which the payment or benefit will be received. The exponent is negative (-*n*) because the potential benefit from present decision-making (i.e., whether to spend or save) will be received in the future. Therefore, Equation 1 is used to calculate the present value (i.e., discounted value) of a financial benefit, adjusted according to the opportunity cost (*r*) related to the delay in time to the benefit's receipt (-*n*) (Fisher, 1930).

When psychology researchers and behavioural economists (Kahneman & Tversky, 1979; Loewenstein & Thaler, 1989) first began to use discounting methods, they employed an exponential function (Grüne-Yanoff, 2015) described by:

$$V = Ae^{-kD} \tag{2}$$

Where *V* is the discounted value; *A* is the initial value; *k* is the discount rate; mathematically analogous to *r* in Equation 1. *D* in Equation 2 reflects the time delay; mathematically analogous to n in the first equation. That said, despite mathematical parity between Equations 1 and 2, in psychological work, the discount rate denotes the degree to which a reward is devalued over time, rather than in economics, where the discount rate reflects opportunity costs. In the equations from both disciplines, the delay is conceptually the same although expressed differently; it is the economist's *n* and the psychology researcher's *D* parameter.

In a typical delay discounting scenario, participants are asked to decide between two options, a smaller-sooner amount of money and a larger-later amount of money. The time (e.g., delay) is systematically manipulated in a series of further questions, allowing researchers to empirically describe the extent to which a sooner (i.e., less delayed) reward is preferred. The outcome being discounted in this example is the monetary amount and the procedure is best described as binary-choice discounting. Binary-choice procedures are when participants respond to the general prompt "Would you rather this (smaller-sooner) or that (larger-later)?" A series of prompts like these allow researchers to calculate an indifference point for each participant at each unit of delay (Bickel et al., 1999). That is, a participant might demonstrate that they consider $200 in a month to be equal (i.e., indifferent) in value to $50 today.

Another discounting procedure used commonly is a fill-in-the-blank style of question (Weatherly & Derenne, 2011). Fill-in-the-blank questions are those where a participant is asked what value of money (received in 1 month) they would consider equivalent to $50 received immediately (i.e., today). Theoretically, both the binary-choice and the fill-in-the-blank processes would find the same indifference points for a participant. In the example given, this would mean that the participant was indifferent between the options of receiving $50 immediately as compared to waiting 1 month to receive $200. Therefore, their indifference point at 1 month of delay is $200.

Researchers have also made use of a visual analogue scale (VAS; Kaplan et al., 2014) to measure the indifference points of participants, which often increases the speed at which participants can complete discounting trials. Similar to the fill-in-the-blank procedure, the VAS means that participants need only respond to one prompt at each delay for researchers to identify their indifference point at that duration of delay. Typically, a VAS ranges from 0-100 left-to-right and includes anchoring statements at the extreme ends of the scale. For example, 0 indicates a very low value and 100 a very high value. In monetary scenarios, the ends of a VAS may be labelled with $0 and $100, for example. A VAS has also been used to obtain a likelihood valuation (e.g., Hayashi et al., 2016; Kaplan et al., 2015; Sargisson & Schöner, 2020), in which case the left-most label could be 0% or “no chance”, while the right-most is 100% or “every chance”.

Discounting methods have been successfully applied to a range of situations requiring that a decision be made. Often (though not always) these applications relate to impulsive or risk-laden decisions. For example, the tendency to text while driving (Hayashi et al., 2016), to smoke cigarettes (Reynolds, 2004), or to engage in environmentally costly behaviour (Carson & Tran, 2009). In each of these examples, an individual must decide whether the immediate outcome is of high enough value or if they would risk choosing a larger or less certain negative outcome.

Reynolds (2004) used a titrating binary-choice procedure to establish whether adult smokers, adolescent smokers, or adult-non-smokers demonstrated the greatest delay discounting of monetary reward values as a function of delay. The indifference points relating to participants' valuation of the initial reward (e.g., $2; immediate certain benefit) compared to a $10 reward received at delays of up to 365 days (delayed, uncertain benefit) were calculated and plotted. Reynolds (2004) showed that adult smokers made the most impulsive decisions ($k = .075$) and that adolescent smokers were comparatively less impulsive ($k = .016$), contrary to the more common trend that adolescents are more prone to impulsivity than are adults (Rosenbaum & Hartley, 2018; Water et al., 2014).

Hayashi et al. (2016) calculated delay discount rates when outcomes were monetary for participants who reported that they often text while driving, compared to those who said they did not text while driving. Participants were presented with a hypothetical scenario where they received a text message while driving and were asked how likely it was that they would reply to the text message given different wait times (i.e., delays). The small and certain outcome was the continuation of a quick and fluid conversation over text-message (i.e., responding to texts quickly; smaller-certain outcome). The risk in the context of the driving scenario was that on any particular occasion when a driver attends to texting and not on their driving, the environment around them could change in such a way that demands their focus to avoid an accident (larger-uncertain outcome). For example, if someone else suddenly stops in the road, texting rather than attending to the road could cause an accident. On the other hand, if nothing unexpected happens and the driving conditions are stable, then perhaps no accident occurs regardless of whether someone is texting and driving. Furthermore, when the driving conditions are poor it may be that a driver's risk calculation is changed. That is, they could discount the value of responding quickly (i.e., with a shorter delay) to the text message in favour of reducing the risk of an accident if, for example, they were driving during a storm.

The risk of an accident was always a part of the decision-making calculus, yet in the latter case the likelihood of an accident is perceptually greater due to the storm and so a delay in responding to one's text messages seem to be more acceptable.

For many of the applications in which psychology researchers utilise discounting, an exponential function provides a poor fit to decision-making data (Myerson & Green, 1995). Indeed, the data from neither of the previous examples (Hayahi et al., 2016; Reynolds, 2004) were well-fit by the exponential function (e.g., Equation 1 and 2). Primarily this is because use of an exponential function relies on the assumption that the rate of discounting is constant for any pair of delays that are equal in duration (Critchfield & Kollins, 2001; Green & Myerson, 1996). Consider the change in opportunity cost (economics) or reward valuation (psychology) between a delay of today and tomorrow, compared to the likely change between 5-years and 5-years-and-1-day from now. In the context of how people value outcomes as a function of delay, typically, there is a greater difference between today and tomorrow than there would be for two days that are equally distant from each other (i.e., 1 day apart), yet are both equally distant from the present (Mazur, 1987).

Hyperbolic Discounting

Compared to exponential models, hyperbola tend to better explain the variance in group- and individual-level discounting datasets in large part because there is no assumption that discounting occurs at a constant rate (e.g., Arnold, 2012; Holt et al., 2003; Myerson & Green, 1995). The formula for a hyperbola has been expressed by Mazur (1987) as:

$$V = \frac{A}{(1+kD)} \tag{3}$$

Where for each participant, V represents their subjective valuation of the absolute outcome amount (A, i.e., initial value), given varied lengths of delay (D) to experiencing that outcome. In hyperbolic delay discounting, the k parameter is a measure of someone's propensity to discount outcomes as a function of delay. Thus, the denominator 1+kD describes the decreasing value of the delayed outcome over time, as people typically become less willing to wait for an outcome they will not experience until a future time (Holt et al., 2003).

Hyperboloid Discounting

Hyperboloid models are the third relevant discounting model. In general, hyperboloid models are based on the same principle as hyperbolic models; discounting rates become smaller as the delay period increases and are not constant (Myerson & Green, 1995). Put another way, if the delay to receiving a monetary reward is already well into the future, an additional day's wait may not impact an individual's valuation of the reward in the same way that a delay of today compared to tomorrow could. Additionally, hyperboloid models extend the form of hyperbolic models (e.g., Equation 3, Mazur, 1987) by the inclusion of an additional parameter.

Two hyperboloid functions have been proposed by discounting researchers. Firstly, Myerson and Green (1995) proposed the following equation:

$$V = \frac{A}{(1+kD)^s} \tag{4}$$

Myerson and Green's (1995) formula works in the same way as Mazur's (1987), with the addition that the whole denominator (i.e., 1 + *kD*) is raised to the power *s*. The *s* parameter is variously described as an individual's sensitivity to delay (Rachlin, 1989), as a scaling parameter (McKerchar et al., 2010; Olsen et al., 2018), and as a constant that determines the discount function shape (Locey & Rachlin, 2015). Secondly, Rachlin (2006) proposed:

$$V = \frac{A}{1+kD^s} \tag{5}$$

Rachlin's (2006) equation is identical to Myerson and Green's (1995) but for the absence of brackets, which alters the order of operations such that *D* is raised to the power of *s* prior to the calculation of *k*.

Equations 4 and 5 both contain two parameters that are free to vary, whereas earlier equations included a single free parameter. As more flexible equations, hyperboloids tend to fit a wider range of discounting data than one-parameter formulae are able to. Furthermore, Equation 5 (Rachlin, 2006) allows for a variable discount rate that can increase as the delay increases. Put another way, Equation 5 tends to better describe patterns of decision-making because it captures changes in the perceived value of an outcome as a function of delay while additionally accounting for changes in the rate of discounting, also as a function of delay (Rachlin, 2006).

Probability Discounting

Similar to data from delay discounting, probability discounting data have been well-fit by hyperbolic (Equation 3) and hyperboloid models (Equations 4 and 5). Mathematically, the hyperbolic model in Equation 3 is the same for describing probability discounting with two exceptions. Firstly, in probability discounting the delay parameter (i.e., *D*) is replaced with the relevant odds-against ratio.

$$\theta = \frac{(1-p)}{p} \tag{6}$$

The second difference when applying discounting equations to probabilistic data rather than delay data is that the parameter used in the former is h, rather than k as is the case in delay discounting (Chung et al., 2021). Neither of these formula adjustments impacts upon the calculation of discounting rates. I give Equation 7 as an example of the minimal changes between, in this case, Rachlin's (2006) formula used for delay data (i.e., as in Equation 5) compared to for use with probabilistic data.

$$V = \frac{A}{1+h\theta^s} \tag{7}$$

Probability discounting procedures have been found useful across several topics including alcohol use (Richards et al., 1999), environmental outcomes (Sargisson & Schöner, 2020), and food intake (Robertson & Rasmussen, 2018). Further, probability discounting has been used as a proxy measurement for risky decision-making and compared across several groups (Green et al., 1999; Reynolds et al., 2004). For example, probability discounting has been used to determine whether adolescents are more prone to discount probabilistic outcomes (Olson et al., 2007), or the extent to which substance users do so (Mejía-Cruz et al., 2016).

Area Under the Curve

An alternative to model-fitting and deriving a parameter (i.e., *k* or *h*) to represent discounting data is the computation of the area under the curve (AUC). This is done by plotting the discounting indifference points and then AUC is calculated using the equation:

$$X2 - X1\left[\frac{Y1+Y2}{2}\right] \tag{8}$$

where the Y-axis (the ordinate) represents the proportion of the non-discounted absolute value and the X-axis (the abscissa) shows the proportion of the maximum possible delay (Smith & Hantuala, 2008). As an example, in delay discounting of monetary outcomes, the Y-axis may be \$0 to \$100, while the X-axis ranges from 0 days (i.e., the outcome occurs immediately) up to a delay of 365 days (i.e., 1 year). Using Equation 8, the AUC is a measure of the extent to which discount rates impact the value of an outcome over time. In the example above, the AUC is the impact that an individual's propensity to discount has on their valuation of monetary outcomes (e.g., \$100) as a function of having to wait for that outcome (e.g., a delay of 1 year). The result of Equation 8 is an area ranging from 0.0 to 1.0, which can be depicted graphically to help with interpretation. AUC values closer to 1.0 mean a lower discount rate, as they indicate a flatter curve.

AUC is widely reported within discounting research and is distinctly different from discounting models (i.e., as described by Equations 1-6) in that it is an atheoretical measurement (Myerson et al., 2001). For example, the subjective valuations as calculated using Equations 1 and 2 rely on the idea that value decreases linearly over longer delays, while the product of Equation 3 relies on the theory that value decreases over delays but to a lesser and lesser extent the longer those delays are from the present. AUC, on the hand, does not rely on any such assumptions.

There are a number of factors used by researchers to determine whether discounting models, AUC, or a combination of both are most appropriate for describing particular discounting datasets. The method of analysis can influence the outcomes of research. Therefore, I reserved the first results section in Chapter 4 for an in-depth examination of the utility of each method prior to subsequent data analysis.

Part 4: Application

On the topic of cyberbystander intervention, I sought to combine theories predominantly from the fields of social and personality psychology with those of behaviour analysis, then applied discounting methods to assess how these ideas relate to decision-making behaviour.

Discounting has been widely utilised across a variety of topic areas (Smith & Hantula, 2008). Some areas of discounting research that are particularly relevant to my thesis include cross-cultural studies (Du et al., 2002; Kirby et al., 2022), studies related to problematic internet use (Chun et al., 2021; Wang et al., 2018), and studies of prosocial types of behaviour (Hayashi & Tahmasbi, 2020; Locey & Rachin, 2015; Rachlin, 2016). Cross-cultural studies are salient given the global pervasiveness of SNS as venues for social behaviour. It is, therefore, essential that my research findings can be applied and adapted across cultures; a task that cross-cultural researchers can inform (Kirby et al., 2022). Additionally, work that has focused on problematic internet use is critical to my own work, as bullying and other forms of targeted aversive behaviour online are increasingly common and problematic (Allison & Bussey, 2016; Boer et al., 2021; Kowalksi et al., 2014). As highlighted in an earlier section (Part 2: Supportive Behaviour Online), the behaviour of bystanders plays a pivotal role in the perpetuation or prevention of such behaviours. In particular, the prosocial and defending behaviour of bystanders is important and therefore their responses to aversive behaviour is a crucial area for research focus (Bastiaensens et al., 2014; Luo & Bussey, 2022; Macháčková & Pfetsch, 2016; Ng et al., 2022).

A meta-analysis by Chung et al. (2021) found that individuals who spend a substantial amount of time gaming online overestimated the likelihood of gaining favourable outcomes compared to the control group; thus, the gaming addicts demonstrated steeper probability discounting. That Chung et al. (2021) documented consistent results from several applications of probability discounting to online gaming suggests that the paradigm can be applied to online behaviours. Furthermore, Hayashi and Tahmasbi (2020) successfully applied a related form of discounting (i.e., social discounting) to the topic of cyberbystander intervention.

Certainty that Harm was Intended

In my series of studies, I approached studying cyberbystander behaviour through a novel use of probabilistic discounting. To do so, I used hypothetical scenarios that were informed by the focus group data (described in Chapter 2) and asked participants to rate their likelihood of intervening. As documented by researchers of bullying (of all kinds, e.g., Fischer et al., 2011; Olweus, 1997) and bystander effect researchers (e.g., Darley & Latané, 1968; Dillon & Bushman, 2015; Macháčková et al., 2015), the ambiguity of aversive situations for those who are not directly involved is a key determinant of whether a bystander will intervene. Therefore, I used probability to manipulate the degree of ambiguity present across different study scenarios and trials.

According to the BIM (Latané & Darley, 1970), less ambiguity tends to mean that bystanders are more likely to intervene. In my study, the BIM (Latané & Darley, 1970) finding should mean that the more certainty there is in a scenario (i.e., less ambiguity), the more likely the relevant bystanders are to intervene. In a recent meta-analysis, Mennicke et al.

(2022) found that there was a paucity of research testing Step 2 of the BIM, which involves recognising a situation as requiring intervention (Latané & Darley, 1968; Latané & Nida, 1981). In particular, I framed the ambiguity around the aggressor's intentions as the *certainty that the aggressor intended to cause harm* (hereafter referred to as certainty). That is, the more certain that bystanders are that harm caused was intentional (i.e., a higher certainty value), the less ambiguous that aspect of the scenario is perceived to be. I used certainty as a probabilistic scale variable such that it was manipulable within a discounting procedure (i.e., an independent variable).

Using certainty as the independent variable and the likelihood of intervening as the dependent variable, the position and form of the curve that best fits the data provide a measure of the participant's perceived value of intervening. The gradient of a discounting curve describes the extent to which a participant's initial preference (i.e., when quite certain, 95%) is devalued as a function of increasing scenario ambiguity (i.e., less certainty). Just as those who show steep delay discounting struggle to delay gratification (i.e., an outcome), steep probability discounting in this context shows that uncertainty about the nature of an online social interaction disproportionately reduces bystanders' intention to intervene.

For the purpose of my probability discounting analyses, I used Equation 6 and converted the certainty values into odds-against ratios. From most certain (i.e., 95% certain that harm was intended) to least certain (i.e., 5% certain that harm was intended) the odds-against ratios were: 0.053, 0.111, 1.000, 2.333, 9.000, and 19.000. In percentage terms, the odds-against ratios are 95%, 90%, 50%, 30%, 10%, and 5%. Using odds-against ratios ensures that higher values (e.g., 19.000 rather than 5%) represent outcomes that are typically the least preferred (and most discounted), such that probabilistic results align with those from delay discounting, where higher values (e.g., 365 days) equate to outcomes that are more delayed and therefore typically less preferred (e.g., Critchfield & Kollins, 2001). Put another way, in my study scenarios when the aggressor's intentions were the most ambiguous (i.e., least certain; 5%), the odds-against value was correspondingly high ($\theta = 19.000$).

Fulfilment of the BIM Steps

Alongside my analysis of cyberbystander behaviour as a function of ambiguity, I also sought to measure the proportion of participants who can be said to have fulfilled individual steps of the BIM. In the bystander effect literature, there is a wealth of evidence supporting that the presence of more than one bystander impacts the likelihood of bystander intervention occuring. However, the baseline prevalence of intervention is often less clearly reported (e.g., McDonald et al., 2016; Van Cleemput et al., 2014), while the fulfilment rate at each step of the model is rarely included at all. In some cases it is difficult to ascertain the baseline prevalence of intervention because the primary dependent measure used is time-elapsed, rather than whether or not intervention occurs (e.g., Fischer et al., 2011; Latané & Rodin, 1969). In others, only sub-scale descriptive statistics were reported and individual survey items measuring types of bystander intervention were collasped in with items less conducive for measuring prevalence (e.g., Ferreria et al, 2020). Still others used Likert measures that were frequency-based where the lowest possible rating represented "almost never", which implies that even the participants selecting this option did intervene some of the time (e.g., Parris et al., 2020).

Due to the inclarity of bystander intervention prevalence rates, as well as the mixed results among those that are reported, I used a threshold of 50% to assess whether my participants fulfilled individual BIM steps. I based the 50% threshold on the prevalence rates reported by several researchers. Latané and Dabbs (1975) reported on the likelihood that a male victim would receive help from a bystander (43%) compared to a female victim (67%). In the absence of information regarding the gender identities of the bystanders who assisted, Latané and Dabbs' (1975) work suggests a prevalence rate somewhere between 43% and 67% (55% if all things are equal). The work of several other researchers further suggests that the prevalence rate of intervention may be approximately 50%; Latané and Rodin (1969) observed that 52% of participants intervened in a single-person robbery and Fischer et al. (2006) found that 50% of their sample intervened in low danger (and 44% in high danger) scenarios if they were a lone bystander.Consequence Type

In addition to the level of certainty (i.e., ambiguity), I manipulated the type and the magnitude of the consequences outlined in each hypothetical scenario. In a pilot study (Chapter 4), I first tested whether social-emotional outcomes produced discounting similar to studies where commodities such as money were used.

Social reinforcers and punishers are critical drivers of behaviour (Guerin, 1994; 2020). Researchers have observed that the work rate of participants incentivised by monetary rewards is indistinguishable from when rewards were social (e.g., Schofield et al., 2015). Furthermore, some behaviours for which the onset is otherwise difficult to explain are products of social reinforcement. For example, starting to smoke cigarettes is often a product of social reinforcers even if the maintenance of smoking is more complex (Biglan & Lichtenstein, 1994; Kobus, 2003). Bullying and other forms of aversive behaviour also tend to be maintained by social reinforcement or punishment (Salmivalli, 2002). Therefore, to best understand cyberbystander behaviour – a response to aversive behaviour – the use of social outcomes was essential for my studies.

Chapter 4: Pilot Study

Outline

In Chapter 4, I progress an argument for the use of discounting methods to better understand the behavioural intentions of cyberbystanders who witness socially aversive behaviour. I present a pilot study (Study 4.1) designed principally to test whether social-emotional outcomes were discounted in a way that is similar to the discounting of money (i.e., a commodity-based outcome). I use the data from this pilot study to examine key questions relating to how researchers approach the analysis of discounting data. Finally, I use those results to outline and justify an a priori plan for the analysis of my future studies.

Probabilistic discounting has been applied to various topics, such as alcohol use (Richards et al., 1999), environmental outcomes (Sargisson & Schöner, 2020), and food intake (Robertson & Rasmussen, 2018). For my purposes, it was critical that I verified probabilistic discounting measures as appropriate when the outcomes are social reinforcers, particularly when applied to social outcomes for bystanders that occur in online spaces, as researchers have demonstrated with social discounting (Hayashi & Tahmasbi, 2020). A distinct strength of discounting methods over other forms of measuring cyberbystander intentions is that, for each scenario, participants provide several responses across trials while one within-scenario variable is manipulated (e.g., certainty). The multiple responses for each scenario measure how a participant would respond more accurately than other forms of self-report (an argument outlined by Fisher et al., 2016), which tend to be single-item measures. If discounting models accurately describe cyberbystander behaviour, it should be possible to study the impact of different contextual factors – and their magnitude - on cyberbystander behaviour. For example, it should be possible to examine the behavioural impact of varying levels of situational ambiguity (i.e., the “certainty” with which a scenario can be interpreted) concerning an aggressor's intentions (Soloman et al., 1982), using the same scenario to retain experimental control.

In Study 4.1, I presented participants with hypothetical scenarios where, across several trials, I manipulated the probability that the aggressor in each scenario intended to cause harm to their target, as perceived by the cyberbystander. For example, in some trials I informed participants (i.e., the cyberbystander) that they were 95% certain that the aggressor had intended to cause harm; therefore, the trial was low in situational ambiguity. In this way, I was able to measure how valuable participants perceived intervening to be and how that perception may change as a function of probability (as a measure of situational ambiguity; e.g., Phares & Wilson, 1972; Solomon et al., 1982). Steep probability discounting would indicate that increasing uncertainty (i.e., greater ambiguity) about the nature of an online social interaction disproportionately reduces bystanders' intention to intervene, similar to how steep delay discounting indicates an inability to delay gratification (Myerson & Green, 1995). The first research question was:

> RQ4.1: Can probabilistic discounting describe cyberbystander decisions to intervene in scenarios with varying levels of certainty regarding whether an aggressor intended to cause harm?

Discounting Research Methods

Between- vs Within-Subjects Discounting Studies

In Part 1 of this chapter, I reviewed methods and analytic approaches for cyberbystander discounting data. I used a within-subject approach because within-subject study designs can allow researchers to compare the same participants' behaviour in different conditions, providing greater control over otherwise confounding variables such as individual biases (Charness et al., 2012; Greenwald, 1976). For my research, this means that if a participant held strong views for or against intervening on SNS, those views will have equally informed their responses across study conditions. Moreover, a within-subjects design does not necessarily preclude between-subject comparisons (Greenwald, 1976), which I also used to explore possible interactions between cyberbystander behaviour and gender identity.

Johnson & Bickel (2008) Criteria Nonsystematic Discounting

Much of the extant literature regarding discounting has included comments on the prevalence and impact of nonsystematic data (e.g., Gilroy et al., 2022; Olsen, 2020; Schlosnagle, 2011). Johnson and Bickel (2008) established guidelines for nonsystematic discounting which consisted of identifying

> (1) if any indifference point was greater than the preceding indifference point by a magnitude greater than 20% of the larger later reward; and, (2) if the last indifference

point was not less than the first indifference point by at least a magnitude equal to 10% of the larger later reward.” (Johnson & Bickel, 2008, p.267, para. 2).

Johnson and Bickel (2008) were careful to state that these guidelines were not necessarily final and complete grounds for data exclusions. Regardless of their caution, these guidelines tend to be more often treated as a rule rather than a starting point from which to adapt (Smith et al., 2018). Using the discounting data from my pilot study (i.e., Study 4.1), I sought to answer:

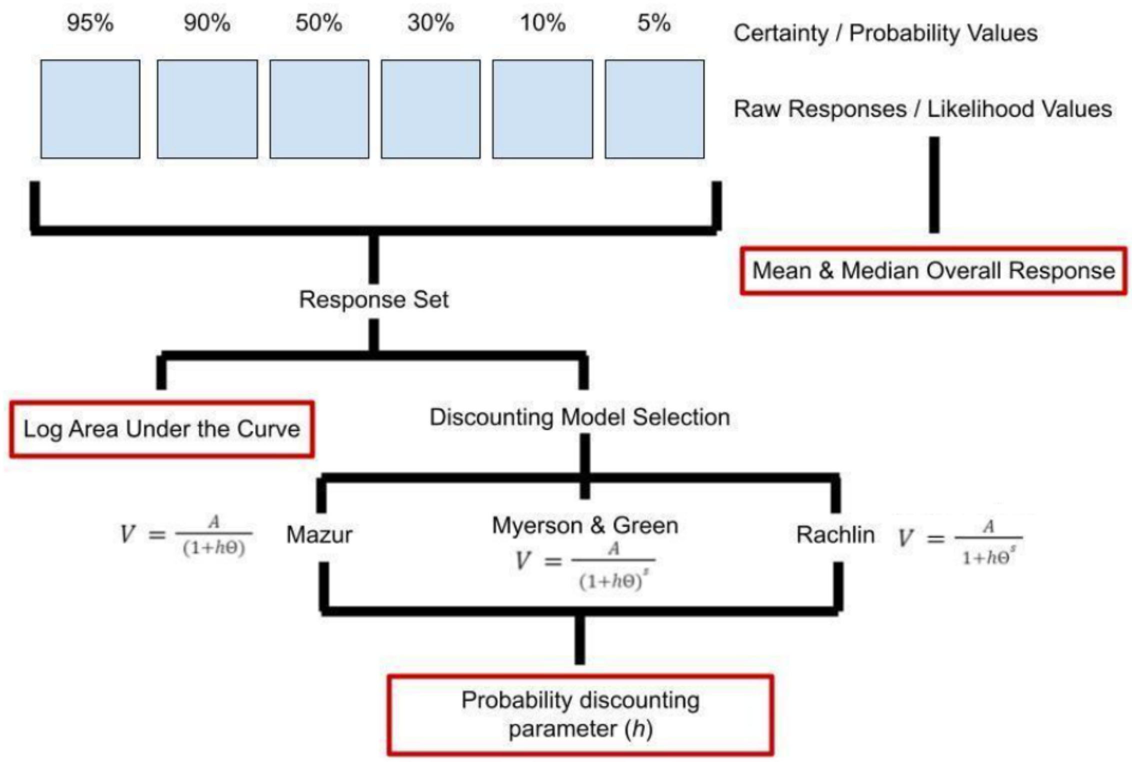
RQ4.2: To what extent are the original Johnson and Bickel (2008) nonsystematic discounting guidelines appropriate for my research, and what adaptions to these guidelines are justified within the novel context of cyberbystander research using probability discounting?

Dependent Variables

An important consideration when conducting discounting research is what metric to use as the dependent variable. I include Figure 2 to show the different dependent variable options for analysing discounting data, as well as how closely these are related to the raw data. Analysis of the raw responses, otherwise referred to as the likelihood of intervening values, is one option (e.g., Belisle et al., 2020). However, comparing data from each of the six levels of certainty and across each of the four study conditions is the extent of testing for which there was sufficient statistical power (Faul et al., 2007). A second option is to calculate the average or median likelihood-to-intervene for each response set. As shown in Figure 2, I use response set to mean the grouping of a participant’s responses within each study condition, e.g., their six responses within the monetary-consequence, high-severity condition. A drawback of using a measure of response-set central tendency (i.e., the mean across the six probability values), though, is that the range of responses as they relate to varying degrees of situational ambiguity (i.e., probability values) are no longer captured.

Figure 2

Flowchart of the Approaches to Data Analysis



More common is the use of metrics that condense the raw data into one or two numbers – as with an average – yet also retain nuance such as how responses might change as a function of probability (Rachlin et al., 1991). Discounting parameters are one way to achieve this (Myerson et al., 2001).

Discounting Model-based Parameters

Various discounting equations and resultant parameters have been used as dependent measurements (Borges et al., 2016). In the previous chapter, I introduced the most used discounting equations: Mazur's (1987) one-parameter equation (EQ3) and the two-parameter models of Myerson and Green (1995; EQ4) and Rachlin (2006; EQ5). Equations with two free parameters like Equations 4 and 5 are more flexible and, therefore, oftentimes provide a better fit to data (McKerchar et al., 2009; Peters et al., 2012).

In probability discounting, the *h* parameter tends to be the primary dependent variable (Rachlin et al., 1991). In my research, the larger an h parameter is, the greater the extent to which someone can be said to devalue probabilistic outcomes (i.e., how far from certain it is that the aggressor intended to cause harm). Put another way, participant intention-to-intervene responses that are well-described using large h parameters decrease more steeply than those those described by smaller h parameters, even when there are only small changes to the level of certainty. For datasets that do not meet Johnson and Bickel's (2008) discounting guidelines, though, the h parameter often takes on an unusually high value (e.g., Smith et al., 2018). Unusually high or low values are an issue when conducting statistical tests because they skew the sample-level distribution of parameter values, rendering them inappropriate for parametric analyses (Field, 2013). High values are further problematic when specifically analysing discounting data because, to accord with theory, parameter values should be bounded (approximately 1.00 and < 20; e.g., Andrade & Petry, 2012; Jarmolowicz et al., 2012; Rachlin, 2006) – though sometimes they are not (e.g., Jarmolowicz et al., 2018; Kaplan et al., 2014; Sargisson et al., 2021).

In addition to *h*, the sensitivity parameter (i.e., *s*) is an important component of many models of decision-making (e.g., Myerson & Green, 1995; Rachlin, 2006). Wakker (2010, p. 6) said that the purpose of s is to account for "the nonlinear ways in which people may process probabilities" and s is just as relevant for describing decision-making as people's processing of the relevant outcomes (as measured by *h* in Study 4.1). In cases where *s* is left to vary, the analysis of the primary discounting parameter (i.e., *h*) becomes complicated (Vincent & Stewart, 2020). To clarify discounting results, the s parameter should be included – but often is not – as a covariate in analyses of the rate of discounting (e.g., *h*). Rachlin (2006) describes s as "dimensionless" if isolated from the respective discounting parameter. Put another way, *s* is interdependent with the *h* parameter. Rachlin also posited that future work in the field of discounting could discern precisely what features of a decision individual participants are said to be "sensitive" to – a research direction that first requires that *s* be fully incorporated into discounting analyses. In still other cases, an alternative approach is to fix the value of the *s* parameter across participants. However, unless the population of interest is relatively homogenous in their sensitivity to the independent discounting variable (e.g., probability in the form of certainty values in my case), this process can negate the model-fitting benefits of using a two-parameter model in the first place (Stewart et al., 2018).

I used the data collected in Study 4.1 to explore the *s* parameter.

RQ4.3: If using model-based parameters, what is the most useful approach to analysing the sensitivity (i.e., *s*) parameter?

Area Under the Curve

An aggregated AUC value is another useful metric for probabilistic discounting (Borges et al., 2016). In the context of my studies, the AUC value represents a participant's overall likelihood of intervening as a function of the situational ambiguity of the scenario (i.e., how certain it was that the aggressor intended to cause harm). As with a shallow discounting curve, a high AUC value generally indicates a high likelihood of intervening. I used a logarithmic transformation (log base 10) on the raw AUC data to account for the non-linear probability scale often used in probability discounting (e.g., Kaplan et al., 2014; Rinderhagen & Sargisson, 2021). Importantly, using a log transformation with a base of 10 also corrects for the fact that, otherwise, the two most certain values of the probability scale (i.e., 95 and 90%) would outweigh the contributions of all other certainty values (Borges et al., 2016; Yoon et al., 2017). Unless otherwise noted, I use the term AUC to mean the Log10 AUC metric.

In sum, the probability discounting parameters *h* and *s* and the AUC metric are both used in the analysis of discounting data, and each has strengths (Myerson et al., 2001; Smith & Hantula, 2008). The type of metric used can critically influence the analysis and interpretation of discounting datasets (e.g., Chung et al., 2021). I used the Study 4.1 pilot dataset to investigate the question:

RQ4.4: Which metric is most appropriate for use as the main dependent variable for my application of discounting methods to the topic of cyberbystander research?

Pilot Study Overview

I used four different scenarios, each briefly depicting a text-based online interaction of the sort that could be found on various popular social media platforms. I tested whether the likelihood of a cyberbystander intervening is related to the degree to which they are certain of the aggressor's intentions (H4.1). Two of the four scenarios involved a monetary consequence while the other two referenced a social consequence, where I implied that failure to intervene could result in that particular consequence. I modelled both the social consequence scenarios on personal anecdotes that were shared in the exploratory focus group phase of my research (Chapter 2). There is a wealth of work from other researchers using similar procedures and various monetary consequences (see Smith et al., 2018, for an overview). Therefore, I included both monetary and social consequences to test that social consequences were discounted similarly to monetary ones (H4.2). Finally, I included low- and high-severity versions of both consequence types as an initial test (H4.3) of the idea that as severity increases, so too does the likelihood of someone intervening to provide help (Bastiaensens et al., 2014). As the primary purpose of Study 4.1 was to establish the suitability of cyberbystander-related social outcomes for use in probabilistic discounting research, I hypothesised that:

H4.1: Participant responses to monetary consequences will be similar to their responses to social-emotional consequences.

I also tested two hypotheses critical to understanding how cyberbystander behaviour functions: H4.2, which I re-test in both Studies 5.1 and 5.2, and H4.3, which is re-tested using data from Study 5.1.

H4.2: The likelihood of intervening will increase the more certain it is that the aggressor intended to cause harm.

H4.3: Participants will rate intervention as more likely when the scenario severity is high.

At the broadest level, I ran a pilot study (Study 4.1) to examine RQ4.5 and subsequently explore variables that may correlate with cyberbystander helping behaviour (RQ4.6).

RQ4.5: Are discounting methods a useful tool for those researching bystander behaviour as it functions in the online environment?

RQ4.6: In what ways (if any) is intention to intervene (including whether intervention utility is discounted) related to other variables (e.g., time online, age)?

Methods

Recruitment

For Study 4.1, I recruited participants using Prolific (www.prolific.co; 2020), an online platform designed to enable the crowdsourcing of genuine research participants. The use of online crowdsourcing platforms to recruit participants has been criticised as increasing the chance of receiving data of poor quality (Brühlmann et al., 2020; Chmielewski & Kucker, 2020). For example, there is reliable evidence that some participants engage in multi-tasking while taking part in online studies, which likely reduces how closely those participants attend to study manipulations (Chandler et al., 2014; Necka et al., 2016). In regard to data quality for crowdsourced discounting data specifically, Craft et al. (2022b) re-analysed a dataset (Craft et al., 2022a) from participants recruited on another oft-used online recruitment platform. They compared the discounting data of those who were included in the original study, those excluded on the basis of poor data quality, and a set of randomly generated data. The excluded data were dissimilar in quality from the included data but similar to the randomly generated data, suggesting that the data quality issue outlined in other research areas is also present in discounting research (Craft et al., 2022b).

To check for inattentive responses in my study, I followed Prolific's guidance (as of 2020) in developing fair manipulation checks (described more fully in the Procedure sub-section). Moreover, among the crowdsourcing options available, Prolific tends to outperform others in terms of participant attention, honesty, and accuracy (Peer et al., 2021; Uittenhove et al., 2023). I applied two screening criteria for Prolific users: (1) the participants should fall within the age range of 16 to 65 years old, and (2) that participants use the internet a minimum of twice per week. All individuals who completed my questionnaire - including those ultimately excluded from the dataset - were paid as compensation for their time at the standard per-hour rate ($6.50 USD; Prolific, 2020). I collected data for Study 4.1 in October and November 2020.

Sample

The questionnaire for Study 4.1 was begun by 204 individuals. Of these, seven began but did not complete it, and I removed their partial data from the final dataset. Of the remaining 197 respondents, a further seven took either too little time (<5mins) or too long (45mins+) to complete the questionnaire. One additional participant answered fewer than 75% of the manipulation check questions correctly. The final sample size was, therefore, 189 individuals.

Most participants were in their mid-to-late twenties (*M* = 27.38; *SD* = 9.63; *Mdn* = 24). The average self-reported time online per day (TO in minutes; *M* = 54.22; *SD* = 26.20; *Mdn* = 49) and time spent on social media (ToSM in minutes; *M* = 24.28; *SD* = 20.53; *Mdn* =

19) suggest that participants were frequent users of online social spaces. My sample included more individuals who identified as male (n = 112) than female (n = 72), as well as five participants who chose other options (*Transgender* = 1; *Not Listed* = 2; *Prefer Not to Answer* = 2).

Apparatus

I recruited all the participants who completed the Study 4.1 questionnaire using the Prolific platform. Prospective participants could read about my study on Prolific (www.prolific.co) before consenting to participate. Participants could complete my Qualtrics-based questionnaire on any internet-capable device. It was important that my questionnaire be easily completed on a smartphone, as it is likely that most participants use SNS on their smartphones more so than on other types of devices (Chun et al., 2020; Reeve et al., 2020) and I wanted participants to take part in my study under circumstances similar to the way they would interact with scenes online that could require their intervention.

Procedure

Discounting

First in every questionnaire were the 24 discounting trials (6-x-4). In each of the four study conditions, I presented a short description of an aversive scenario that could be encountered on social networking sites (SNSs). To maintain consistency across conditions, I kept the length and language in each scenario as similar as possible. For each of the six trials within each condition, I asked participants to imagine that they had varying levels of certainty in whether the aggressor was intending to cause harm. In line with the probability discounting procedures used by other researchers (e.g., Kaplan et al., 2014; Rinderhagen & Sargisson, 2021), I asked that participants imagine that they were either 95%, 90%, 50%, 30%, 10%, or 5% certain. The levels of certainty were presented in a randomised order, and every trial in the discounting section concluded with the same general prompt question: “How likely are you to intervene?”

Participants responded to each “How likely” prompt by moving an indicator on a visual analogue scale (VAS) that ranged from 0%, labelled as “Unlikely to reply,” through to 100%, “Will definitely reply” (e.g., Brumfitt & Sheeran, 1999; Kaplan et al., 2014). For each trial, I set the default position of the VAS marker to the midpoint (i.e., 50%) of the scale and the Qualtrics question validation to require at least one click to minimise the chance of satisficing. Compared to other response procedures, responses using a VAS tend to be fast and retain measurement validity (Hayes, 1921; Jenkins et al., 2020; Valle-Jones et al., 1992). As discounting procedures can become repetitive for participants (Mitchell, 1999; Smith & Hantula, 2008), I considered the time saved by using the VAS rather than a titrating procedure to be a critical factor for the validity of my dependent measurement (i.e., the likelihood of intervention).

In this pilot study, Study 4.1, I manipulated the consequence type (monetary or social-emotional) and the scenario severity (low- or high-severity scenario). The scenario with a monetary consequence that was high in severity (MCH) read:

> Imagine that you have heard through colleagues that an online scam is going around. The topic comes up in a group conversation and you learn that people are targeted via their public social media profiles, in a very specific way that you all discuss in detail. It is less clear how they scam their targets out of money, though you have heard of some people losing as much as $500. A few days later you see a social media post that is directed at a friend of yours.

The low-severity scenario with a monetary consequence (MCL) was:

> Imagine that one of your friends has been struggling to promote their on-the-side business creating unique craft projects to order. Your friend sells their work for a

> profit, usually between $20-$50. Recently they have been using social media to raise the profile of the business, with a particular focus on displaying reviews from previous customers to establish the quality of their work. You happen to be on social media when a new review is posted publicly.

To compare whether social-emotional consequences function in a similar way to monetary ones, I included this scenario based upon earlier focus group findings (Chapter 2) for when the scenario is high in severity (ECH):

> Imagine that you are online when you come across a nasty rumour about a close friend of yours. A recent ex-partner of theirs has publicly posted on social media, accusing your friend of treating them badly and being wholly responsible for the breakdown of the relationship. They have posted publicly and so the accusations are easily viewable to your friends, as well as all their friends and family.

Then, to compare whether social-emotional consequences function similarly in low-severity scenarios (ECL), I used the following vignette based on an anecdote shared in a focus group:

> Imagine that you see a social media post in which someone appears to have an issue with one of your close friends. Your friend is being accused of being rude for something you know they could not have done. Knowing your friend, they would be embarrassed to know that they had potentially offended someone, let alone that their own friends and family could come across the accusation.

As a manipulation check, I included a multiple-choice question after each block of six trials. For the check, participants needed to identify the topic statement that matched the scenario they had most recently completed. The multiple-choice question always had four response options that described the four different study condition scenarios, thus, was a test of recognition rather than recall. The topic statements were: “A potential $500 scam” for the

MCH scenario, “A friend’s craft business” for the MCL scenario,” “Someone’s ex-partner lashing out” for the ECH scenario, and “Your friend being accused of being rude” rude for the ECL scenario (discounting section information is also presented in Appendix B).

Additional Measures

It is likely that the benefits of social media are not equally distributed between different groups of people. Relatedly, the risks of social media, particularly if intervening in others’ social interactions, also may not be equal across groups of people (Pouwels et al., 2022). To examine potential between-person correlates of cyberbystander behaviour, I included questions regarding participants’ time spent online (TO), time spent on social media particularly (ToSM), the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965), an item measuring perceived social status (PSES; Adler et al., 2000), and demographic items (see Appendix C for further detail). Based on others’ research, these facets may be related to cyberbystanders’ intentions to intervene when they see aversive social behaviour. For example, participants who spend more time on social media (i.e., who measure highly on ToSM) could be confronted with situations calling for cyberbystander intervention more often than others (Barlett et al., 2019; Park et al., 2014), which may mediate their intention to intervene. Additionally, participants who score highly on the RSE and who rate themselves more highly in relation to others in society (PSES) may feel more capable of intervening and, therefore, do so more often (Paulhus & Martin, 1987; Piff et al., 2010; Roche et al., 2002).

Analysis

Prior to collecting data, I completed a power analysis using G*Power v3 (Buchner et al., n.d., https://tinyurl.com/424kbpnz), where I input a range of possible effect sizes in lieu of a directly comparable study. Following psychological research conventions (β = .80; α = .05; assuming a small effect size; Field, 2013), I found that a sample size of at least 128 participants was appropriate for a pilot study using a 2-x-2 repeated-measures ANOVA.

Findings & Discussion

Part 1: Establishing a Plan for Analysis

Can Discounting be Applied? (RQ4.5)

In examining the Study 4.1 data, I found ample support for the use of discounting to better understand cyberbystander behaviour (RQ4.5). I first checked that each of my cyberbystander scenarios were identified correctly in the manipulation check questions. Across the four manipulation checks, most participants correctly identified the relevant scenario (78.31%). Furthermore, the only participants who failed a majority of the manipulation checks were those who I had already excluded from the sample on the basis that they took too little time (< 5 minutes) or too much time (> 45 minutes) to complete my questionnaire. Next, I reviewed the raw discounting data. The skew of the distribution of responses at each certainty value ranged from -2.57 (95% certainty) to 0.42 (5% certainty) but, as expected, the overall response sets were normally distributed (Response Set Average Intention to Intervene: $F(189) = 0.054$, $p = .200$). I use "response set" to mean the group of responses that a participant gave within each study condition, as displayed in the Figure 2. A response set, then, comprised a participant's response to the six certainty-that-harm-was-intended values: 95%, 90%, 50%, 30%, 10%, and 5%.

Nonsystematic Discounting (RQ4.2)

I input each participant's raw response sets into the discounting model selector software (DMS; Gilroy & Hantula, 2018), which among other functions, can indicate whether a response set passes one or both of Johnson and Bickel's (2008) guidelines for determining nonsystematic discounting. Of the 756 total response sets (4 conditions x 189 participants), the DMS showed that 23% did not pass the standard guidelines, thus, could be considered nonsystematic (Johnson & Bickel, 2008). Although 23% of a sample being categorised as nonsystematic may seem like an issue, it is common for between 20 and 30% of a dataset to fail the guidelines (Gilroy et al., 2022; Smith et al., 2018). I conducted a chi-square test to check whether answering the manipulation checks correctly or incorrectly was related to meeting Johnson and Bickel's (2008) data guidelines, finding that these were unrelated, $X^2(1) = 1.52$, $p = .217$, $\Phi = .045$. I conducted a second chi-square test regarding whether the proportion of response sets that met or did not meet the guidelines was related to my study conditions, which they were not, $X^2(3) = 5.24$, $p = .14$, $\Phi = .09$. Furthermore, I tested the AUC values for response sets that met the guidelines and found that there was a chance of underestimating participant's intention to intervene by adhering to the original guidelines (Johnson & Bickel, 2008). In particular, analysis of the MCH responses would be most impacted by the removal of all response sets considered nonsystematic, $t(60.87) = 2.28$, $p = .02$.

I identified several patterns in the data within the response sets that did not meet the original Johnson and Bickel (2008) guidelines, that is, were nonsystematic. Of those nonsystematic data, 56% were nonsystematic due to a pattern of static responding. I defined static responding as cases where all six indifference points fell within a 10% range. Most cases of static responding were when participants rated intervention as highly likely (e.g., a response set consisting of 99%, 95%, 94%, 97%, 93%, and 90%). Approximately 14% of the nonsystematic response sets had a U-shaped pattern (3% of the full dataset). That is, indifference points began high when the certainty was also high (e.g., 99% likely to intervene when the certainty was 95%), then decreased (e.g., 30% likely when the certainty was 50%) only to increase by 10% or more at the lower levels of certainty (e.g., 80% likely when the certainty was 5%). The remaining 41% of nonsystematic response sets (10% of the full dataset) either showed an increase as the ambiguity grew (i.e., less certainty) or showed no clear pattern.

Study 4.1 data broadly met the expectation for discounting data in that the rate of nonsystematic discounting fell within the expected range (Smith et al., 2018), providing an answer to the first part of RQ4.2. Furthermore, the guidelines for what constitutes nonsystematic data should be reviewed and modified as part of applying discounting to novel research contexts (Johnson & Bickel, 2008), although often they are not. For example, Smith et al. (2018) reviewed 114 academic articles in which discounting studies were reviewed, and in only 14 of these, the guidelines were modified. In answer to the latter part of RQ4.2, I did not exclude any response sets on the basis of failing to meet the guidelines proposed by Johnson and Bickel (2008). By retaining what would otherwise be nonsystematic response sets and treating these as important insights into the behaviour of interest (i.e., cyberbystander behaviour) rather than as methodological anomalies (Ballard et al., 2023), my approach was considerably different to the Johnson and Bickel’s guidelines.

Dependent Variable(s) (RQ4.3 & RQ4.4)

A critical question I explored in Study 4.1 was which metric is most appropriate for use as the primary dependent variable in my subsequent studies (i.e., RQ4.4). As in Figure 2, each response set average or median response is an option for use as a dependent variable, and both are closely derived from the raw data. The response set average is useful as a measure of people’s general tendency to intervene (Step 5) or to fulfil earlier steps in the BIM; however, if used alone, neither of these measurements captures any difference in the within-scenario variable. In Study 4.1, this means that the response set mean or median could be used to compare across scenarios (e.g., consequence type and severity level) but not to investigate variations in certainty.

Another option for a dependent variable was the discounting parameters that are the product of finding the best-fitting model of the raw discounting data (see Figure 2). For each response set, I used the discounting model selector (DMS; Gilroy et al., 2017) to ascertain which discounting equation and, subsequently, which *h* and *s* parameter values were the best fit for each response set. Rachlin's (2006) hyperboloid provided an overall good fit for my participants' response sets (Equation 7 best fit 85.05% of all response sets), which provides partial support for the use of discounting in describing cyberbystander decision-making (i.e., RQ4.1). However, there are theory-based concerns regarding the appropriateness of using parameters born from discounting equations (e.g., Smith et al., 2018; Stewart et al., 2018; Vincent & Stewart, 2020).

An alternative to parameters based on model fitting is the use of AUC. A key advantage of AUC is that it is an atheoretical measure – its validity is not tied to how accurately a model describes the raw data (Borges et al., 2016; Myerson et al., 2001). Briefly, AUC is calculated by taking the associated trapezoid of each of the six indifference points (i.e., the raw data for each of the certainty values) and extrapolating from these to account for the missing certainty values. For example, I did not ask participants what they would do at 15% certainty, but the AUC calculation accounts for these missing scale values. Some researchers have found that results based on AUC indicate more aversion to risk than *h* and *s* parameters from the same dataset, which suggests there may be an important discrepancy between the accuracy of the *h* and *s* parameters compared to the AUC values (e.g., Chung et al., 2021).

A distinct advantage of AUC over model-derived parameters is that AUC values are more likely to be normally distributed (Borges et al., 2016). When the distribution of the averages for a series of samples approximates statistical normality, a broader range of analyses are appropriate for use with the dataset (Field, 2013). Using the MCH condition in my study (Study 4.1) as an example, the AUC values were within the acceptable range to meet the assumption of a normally distributed dataset (*Skewness* = -.64; *Kurtosis* = .40). On the other hand, the distribution of the *h* and *s* values derived from fitting Rachlin's (2006) equation to the same data was heavily skewed (h = 11.89; s = 6.98) and leptokurtic (h = 145.64; s = 55.08); this level of deviation from a normal distribution renders all parametric tests inappropriate.

Determining the Primary Dependent Variable

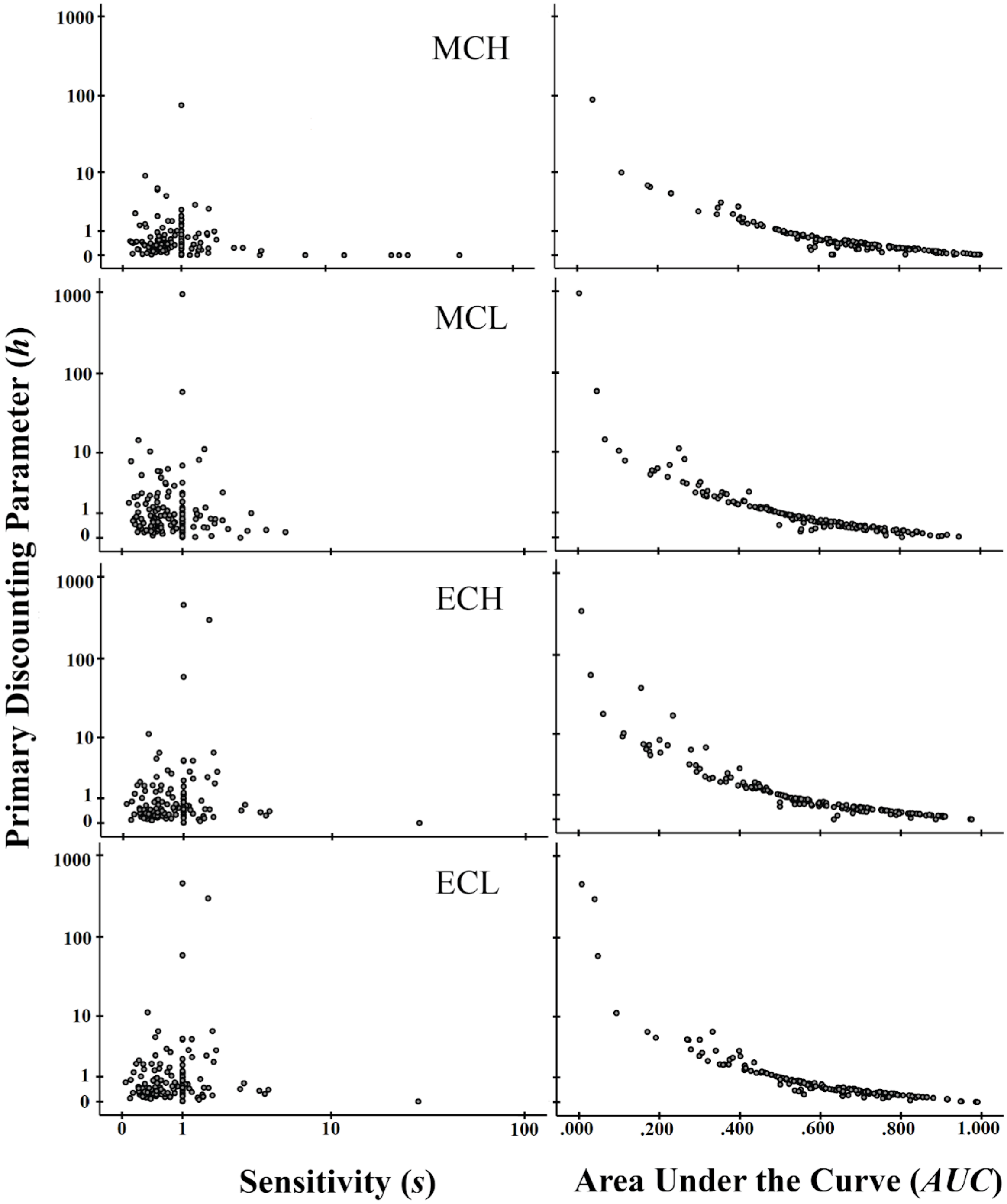
As others have put it, researchers are often comparing measurements against each other to ascertain which is the most useful description of behaviour (Bunce, 2009). Thus, in the case of research using discounting, a crucial question is whether model-derived parameters or the atheoretical AUC measure provide the most accurate description of decision-making. Using a Bland-Altman (Bland & Altman, 1986) chart to inspect two approaches is beneficial because any changes in the differences (i.e., the disagreement between measures) as a result of the magnitude of the measures are apparent. The disagreement between measures is plotted on the y axis ($X_2 – X_1$) as a function of the average of the measures to account for magnitude (Bunce, 2009). For example, h and s parameters might make sense when both are less than one, but when h exceeds 10, they may be less sensical. Furthermore, the technique described by Bland and Altman (1986) is useful for exploring the relationship between the two probability discounting parameters (h and s); thus, is useful regarding RQ4.3.

Smith and Hantula (2008) advise careful consideration of Bland-Altman analyses when making the decision to use h parameters or AUC values for discounting analyses. Their analyses showed that the patterns of results between two tasks (e.g., with a $1,000 vs. a $10,000 outcome) differed markedly depending on which metric was used. Further, some of their main and interaction effects were highly dependent on whether AUC or the h parameter was used. In Smith and Hantula's case, the use of the AUC metric revealed a clear and statistically significant difference between the $1,000 and the $10,000 outcomes, whereas when the *h* parameter derived from the same data was used, the result was non-significant, and the difference between tasks notably reduced (Smith & Hantula, 2008).

In line with Smith and Hantula's (2008) approach, I conducted Bland-Altman analyses of the candidate metrics for use as my primary dependent variable. In Figure 3, I present a panel of eight plots: from top-left clockwise, comparisons of *h* and *s* from the MCH condition, *h* and AUC from the MCH condition, then for each subsequent row the same *h* and *s* (left) and *h* and AUC (right) plots for MCL, ECH, then lastly ECL conditions. All plot axes are in a log10 scale to match the format of the key measures; the *h* and *s* parameters and AUC. The correlation scatterplots (i.e., Figure 3) comparing the *h* parameter with both the *s* parameter and, separately, the AUC value highlight a small number of extreme *h* values. I removed the most extreme *h* values (> 1000) from the Bland-Altman (1986) plots in Figure 3 to examine the congruence between these discounting metrics (*h, s*, and *AUC*). That there are such extreme outlying *h* values was also an issue for subsequent statistical analyses due to the skewed distribution (Smith & Hantula, 2008). In addition, Spearman's correlation tests, presented in Table 4, were inconclusive regarding whether the *h* and *s* parameters correlate. Ordinarily, it would be of benefit for two parameters in a model not to correlate as it suggests that the combination of two independent factors describes the dataset better than a single parameter. However, the *h* parameters calculated using the best-fitting model (i.e., Equation 7; Rachlin, 2006) are contaminated by *s* (p. 3, Vincent & Stewart, 2020). Therefore, the *h* and *s* parameters are not independent, rendering it inappropriate to compare *h* values across participants.

Figure 3

Distribution of Discounting Measures



Note. The left side of the figure are scatterplots of pairs of *h* and *s* parameters; the right side are scatterplots of *h* and *AUC* measures. In order from top-to-bottom are the data from the MCH, MCL, ECH, then the ECL conditions. With one exception (the AUC scale is bounded at 1.000), all axis scales are presented as Log10 to best account for outlying datapoints.

Table 4

Correlation Between Discounting Measures

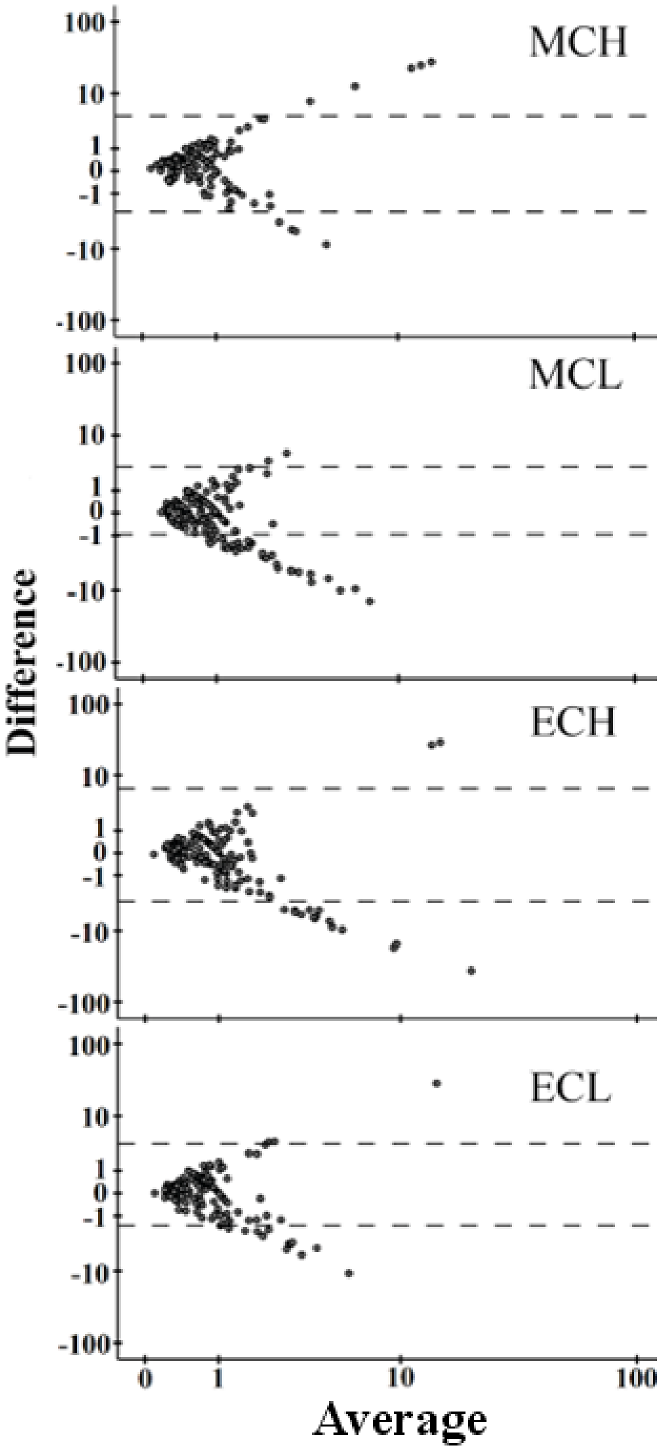
	h vs *s*		*h* vs *AUC*	
	r_s	p	r_s	p
MCH	-0.082	0.307	-0.913	< 0.001
MCL	0.197	0.013	-0.965	< 0.001
ECH	-0.170	0.031	-0.977	< 0.001
ECL	0.197	0.013	-0.965	< 0.001

Note. Correlation coefficients (left columns) and associated significance values (right columns) are from Spearman's analyses as the h and s metrics violate the assumptions of Pearson's test.

As in the previous correlation analyses, Figure 4 also contains eight panels derived from data from the MCH, MCL, ECH, and ECL conditions. Data in the left-hand panels plot pairs of *h* and *s* discounting parameters (*h* on the *y* axis, *s* on the *x* axis), while right-hand plots are of the *h* parameter against the AUC values (*h* on the *y* axis, AUC on the *x* axis). Regardless of the condition, patterns emerge in the *h* versus *s* plots and the *h* versus AUC plots. The difference between each pair of *h* and *s* parameters clusters between zero and one. However, as the value of the parameters increase, the difference increases in a bidirectional fashion. That is, sometimes larger scores deviate such that the *h* value is greater than the *s* value, but sometimes such the *s* value is greater than the *h* value. This result is more instructive than a standard correlation analysis because it shows how the size of the score impacts the degree of difference between the two discounting parameters (Bunce, 2009). That said, the conclusion from both types of measurement analyses is consistent in that the *h* and *s* parameters are inconsistently related, which raises a critical problem because the calculation of *h* is partially dependent on the values of *s*.

Figure 4

Bland-Altman Plots of Discounting Measures



Note. Each panel compares pairs of *h* and *s* parameters; dashed horizontal lines indicate the 95% confidence intervals for the differences between measures. In order from top-to-bottom are the data from the MCH, MCL, ECH, then the ECL conditions.

Taken altogether, the results of these pilot analyses have informed my decision to use AUC as the primary dependent variable in subsequent studies. When compared to each other, the *h* parameter and AUC share a strong, positive correlation which suggests both are able to describe discounting satisfactorily (see Figure 3; Table 4). On the other hand, the relationship between *h* and *s* parameters is much less clear, which further complicates the existing issues with the analysis of these metrics (e.g., Stewart et al., 2018; Vincent & Stewart, 2020). In contrast to others' work (e.g., Chung et al., 2021), in the context of my research, analyses using the *h* and *s* parameters were also more likely to underestimate participants' intentions to intervene compared to AUC. Put another way, an analysis using the AUC metric was also best able to represent the full spectrum of likelihoods of intervention. On the other hand, to use the *h* and *s* parameters would have necessitated the transforming or removal of some sets of participant data, particularly the data from those who indicated a high likelihood of intervention – the participants who are of most interest at this stage of cyberbystander research.

Part 2: Pilot Study Findings & Discussion

Discounting Methods for Cyberbystander Research (RQ4.1, RQ4.5, & H4.2)

As previously outlined in, I use the terminology *raw responses* to refer to each individual data point. For example, a single raw response would be one participant's response, in one study condition, at one level of probability (i.e., P32's answer at 50% in the MCH condition). In Table 5, I have presented all relevant descriptive statistics for each certainty value and each condition as a subset. The highest intention-to-intervene values tended to occur in the MCH study condition, although particularly at high levels of certainty (95% and 90% certain), the data from this condition also deviated most from being normally distributed (kurtosis values of 5.95 and 6.22, respectively). In support of H4.2 at the aggregate level, all conditions resulted in high intention-to-intervene values when the certainty was high (i.e., most certain = 95%), then, that likelihood decreased as a function of growing uncertainty (i.e., most ambiguous = 5%).

Table 5

Descriptive Statistics at each Certainty Value within each Study 4.1 Condition

	95%				90%				50%				30%				10%				5%			
	MCH	MCL	ECH	ECL	MCH	MCL	ECH	ECL	MCH	MCL	ECH	ECL	MCH	MCL	ECH	ECL	MCH	MCL	ECH	ECL	MCH	MCL	ECH	ECL
M	88.69	81.49	82.14	84.48	87.93	79.06	80.90	82.57	69.88	57.60	57.44	60.53	56.78	44.08	43.39	48.31	45.33	31.67	33.61	38.71	39.82	25.67	28.19	33.48
SD	23.16	26.71	27.34	24.14	21.82	26.09	25.87	25.59	25.60	25.57	26.06	25.33	28.28	27.81	28.48	28.25	32.90	27.74	30.39	31.17	34.42	27.92	31.04	32.04
SEM	1.68	1.94	1.99	1.76	1.59	1.89	1.88	1.86	1.82	1.86	1.90	1.84	2.06	2.02	2.07	2.06	2.39	2.02	2.21	2.27	2.50	2.03	2.26	2.33
Mdn	100	95	95	95	100	90	91	92	95	60	60	64	60	40	40	46	40	23	22	30	31	19	17	20
Skew	-2.57	-1.71	-1.79	-2.01	-2.52	-1.51	-1.71	-1.97	-0.85	-0.46	-0.41	-0.66	0.03	0.27	0.30	0.00	0.27	0.77	0.62	0.41	0.42	1.11	0.92	0.65
Kurt	5.95	2.11	2.21	3.51	6.22	1.49	2.18	3.11	0.14	-0.45	-0.52	-0.22	-1.07	-0.78	-0.94	-0.98	-1.25	-0.34	-0.86	-1.10	-1.21	0.36	-0.43	-0.86

Note. 95%, 90%, 50%, 30%, 10%, and 5% are certainty values, *M* = mean indifference point, *SD* = standard deviation, *SEM* = standard error of the mean, *Mdn* = median indifference point, *Skew* = skewness, *Kurt* = kurtosis of the distribution of the indifference points (*N* = 189). MCH = monetary consequence high in severity; MCL = monetary consequence low in severity; ECH = emotional/social consequence high in severity; ECL = emotional/social consequence low in severity.

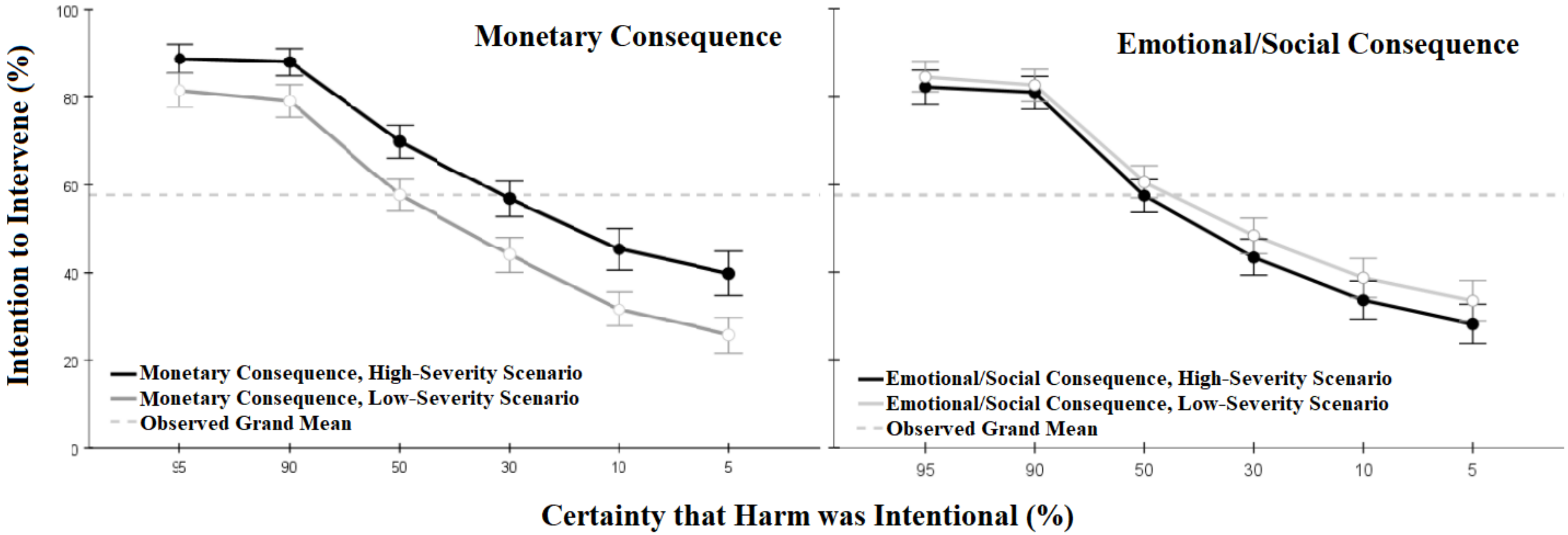
Consequence Type (H4.1) X Severity (H4.3)

I ran a repeated-measures ANOVA using the indifference point data as an initial step in the analysis of the Study 4.1 dataset. The ANOVA had two independent variables: certainty (six levels: 95%, 90%, 50%, 30%, 10%, and 5%) and study condition (four levels: MCH, MCL, ECH, and ECL). The distribution of responses at several of the indifference points deviated from normal; therefore, the ANOVA results should be interpreted cautiously. Further, I used the Greenhouse-Geisser corrected results because Mauchly's test suggested that the assumption of sphericity was not met (Field, 2013), $W_{CERT} = .01$, $X^2(14) = 851.30$, $p < .001$ and $W_{COND} = .90$, $X^2(5) = 19.10$, $p = .002$. Both main effects were statistically significant. That is, responses differed as a function of certainty, $F_{CERT}(1.53, 287.25) = 577.26$, $p < .001$, $R^2 = .75$; and study condition, $F_{COND}(2.81, 528.95) = 18.99$, $p < .001$, $R^2 = .09$. There was no significant interaction between certainty and study condition, which supports H4.1, $F(7.60, 1428.24) = 1.97$, $p = .05$, $R^2 = .01$.

All pairwise comparisons of the certainty levels - with one exception - were statistically significant even when accounting for the full suite of comparisons (Bonferonni correction $\alpha = 0.05 / 21 = .002$). The exception was the responses at 95% and 90% certainty, which did not markedly differ from one another, $M_{Diff} = 1.58$, $p = .008$, 95% CI [.25, 2.92]. Additionally, pairwise post hoc tests of the study-condition variable showed that responses in the MCH condition statistically differed from those in the MCL condition, a comparison displayed in the left panel of Figure 5, $M_{Diff} = 11.48$, $p < .001$, 95% CI [6.79, 16.16]. Responses in the MCH scenario also differed from those in the ECH condition, $M_{Diff}= 10.46$, $p < .001$, 95% CI [5.52, 15.41]. The indifference values from the two emotional-social conditions (ECH and ECL) did not significantly differ, $M_{Diff} = 3.74$, $p = .07$, 95% CI [-.21, 7.68] (right panel of Figure 5), and nor did the MCL condition values differ significantly from the ECL values, $M_{Diff} = -4.75$, $p = .01$, 95% CI [-8.73, -.77].

Figure 5

Discounting of Monetary versus Emotional/Social Consequences



Note. 95% confidence intervals are indicated by the error bars. Results from the conditions with a monetary consequence are in the left panel, those from the emotional/social consequence conditions are in the right panel. The grey lines with open circles = the low-severity conditions. The black lines with closed circles = the high-severity conditions. The dashed grey line indicates the estimated grand mean of all conditions.

I used the DMS (Gilroy & Hantula, 2018) to calculate a logarithmic (base 10) AUC measurement for each participant within each condition. I included descriptive statistics for the AUC measures in Table 6. Of note is that the most common AUC value in all conditions was 1, or 100% likelihood, which means several participants said they would always intervene. Furthermore, the skewness and kurtosis metrics for each condition affirm that, in the aggregate, the data and variance approximate a normal distribution and are suitable for parametric statistical analysis.

Table 6

Area Under the Curve Descriptive Statistics for Each Condition in Study 4.1

	MCH	MCL	ECH	ECL
Mean	0.674	0.556	0.562	0.600
SEM	0.016	0.016	0.017	0.017
SD	0.219	0.220	0.239	0.232
Skew	-0.640	-0.420	-0.376	-0.660
Kurt	0.395	0.091	-0.295	0.238
Median	0.679	0.561	0.558	0.622
Mode	1.000	1.000	1.000	1.000

Note. MCH = monetary consequence, high severity; MCL = monetary consequence, low severity; ECH = emotional/social consequence, high in severity; ECL = emotional/social consequence, low in severity.

I conducted a 2-x-2 repeated-measures ANOVA to test participants' AUC values for an effect of either consequence type (monetary or social-emotional; H4.1), severity level (low- or high-severity; H4.3), or their interaction. There was a significant main effect of consequence type, $F(1, 187) = 4.79$, $p = .03$, $R^2 = .03$, severity level, $F(1, 187) = 8.19$, $p = .01$, $R^2 = .04$, and an interaction effect, $F(1, 187) = 34.10$, $p < .001$, $R^2 = .15$.

Pairwise comparisons with Bonferroni corrections clarified that the notable differences were between the MCH condition and all others ($M_{Diff\ MCH\text{-}MCL} = 0.12$, $p < .001$; $M_{Diff\ MCH\text{-}ECH} = 0.11$, $p < .001$; $M_{Diff\ MCH\text{-}ECL} = 0.07$, $p = .001$), as well as between the two low-severity conditions ($M_{Diff\ MCL\text{-}ECL} = 0.04$, $p = .027$).

In sum, my results support the hypothesis that intervention tended to be more likely when the scenarios were high in severity rather than low when the consequence was monetary (i.e., H4.3). When the consequence was emotional or social, at the group level, the severity level did not have a significant impact on intervention. Moreover, the results from the monetary-consequence conditions were like those of the emotional-social conditions in that both outcomes (i.e., monetary, and social) were discounted in a hyperboloid-like fashion (i.e., H4.1). Ultimately, my analyses suggest that it is appropriate and useful to apply discounting to the study of cyberbystander behaviour (RQ4.5) – although continuing research to understand the strengths and limitations of this application of discounting methods is critical.

Covariates (RQ4.6)

Based on the existing literature (reviewed in Chapter 3), I included several variables that may have correlated with decision-making on the topic of cyberbystander intervention. I present the correlation coefficients and the associated significance values in Table 7, with statistically significant pairs in bold font. The two initial time-related variables (i.e., TO and ToSM) were significantly correlated (Table 7). I calculated TO and ToSM in the same way, specifically: Weekly Average Hours = (Weekday Time x 5) + (Weekend Time x 2). Additionally, I calculated the proportion of online time (PoT) that participants spent on SNS (PoT = TO / ToSM) to explore whether internet availability might have had an impact on SNS use. For example, some participants may have little time available to spend on the internet and so their time spent on SNS is correspondingly low (i.e., participant have a low ToSM score), however, if most of the online time they do have available is spent on SNS this will be reflected in the PoT measure (i.e., the same participant would have a higher PoT score). For Study 4.1 participants, there was a strong positive correlation between the ToSM and PoT measures, $r(187) = .752$, $p < .001$.

Table 7

Correlations Between Additional Measures and the Intention to Intervene AUC Values from Study 4.1

	TO	ToSM	PoT	PSES	RSE	MCH	ECH	MCL	ECL
TO						-0.019	-0.088	-0.054	-0.101
						0.798	0.231	0.458	0.165
ToSM	**0.284**					0.053	-0.009	0.084	-0.036
	< .001					0.467	0.902	0.252	0.624
PoT	**-0.336**	**0.752**				0.058	0.023	0.097	0.034
	< .001	**< .001**				0.428	0.757	0.184	0.638
PSES	-0.046	0.033	0.020			0.093	**0.195**	0.064	0.091
	0.527	0.652	0.780			0.205	**0.007**	0.381	0.213
RSE	0.056	0.122	0.078	**-0.355**		-0.097	-0.090	-0.098	-0.056
	0.447	0.094	0.287	**< .001**		0.186	0.216	0.180	0.441
Age	-0.067	-0.113	-0.091	-0.019	0.005	-0.067	-0.079	**-0.231**	-0.083
	0.358	0.123	0.217	0.799	0.949	0.364	0.280	**0.001**	0.258

Note. TO = time online; ToSM = time on social media; PoT = proportion of time online that is spent on social media; PSES = perceived socioeconomic status; RSE = Rosenberg's Self-Esteem Scale; MCH = monetary consequence, high severity; ECH = emotional/social consequence, high severity; MCL = monetary consequence, low severity; ECL = emotional/social consequence, low severity. All statistics presented are Pearson's correlation coefficients (top) followed by the associated significance value (bottom). Significant values are in bold font.

Further shown in Table 7 is a negative correlation between PSES and RSE scores, $r(187) = -.355$, $p < .001$. There were also two significant correlations between a measure of discounting and an additional measure. First, AUC values from the ECH condition shared a correlation with PSES, such that those who rated themselves as high in socioeconomic status also reported a tendency towards intervening when the stakes were high and social in nature, $r(187) = .195$, $p = .007$. Second, age and likelihood of intervening when the scenario was low in severity with a monetary consequence (MCL condition) were negatively correlated such that younger participants were less likely to intervene, $r(187) = .231$, $p = .001$.

In answer to RQ4.6, intention to intervene may be variably related to PSES (in severe situations with social consequences; ECH) and to age (in less severe situations when consequences are monetary; MCL). As an exploratory research question, further research to clarify the relationships between the discounting of the utility of intervening online and these additional measures is needed.

General Discussion of Chapter 4

Discounting Methods for Cyberbystander Research

I reviewed a range of approaches for analysing discounting data to determine which was most appropriate for the analysis of cyberbystander behaviour. In the existing discounting literature, there is some lack of clarity related to whether different topics of study necessitate adjustments to what are otherwise the “gold standard” discounting methods (e.g., Johnson & Bickel, 2008; Smith & Hantula, 2008), therefore, I addressed this with RQ4.2. Johnson and Bickel’s (2008) guidelines for nonsystematic data have become commonplace in the field of discounting research – despite both authors’ early warnings of a one-size-fits-all strategy. In addition, when there are discrepancies between model-fitted metrics and those more closely derived from the raw data (e.g., the indifference points), the mathematical nuances of the former often seem to take precedence over what may in fact be best explained by conceptual aspects of the study (Smith & Hantula, 2008). For example, if the primary discounting parameter (e.g., *h* for probability discounting) for a participant is unusually high (i.e., greater than 10), a common approach is to remove the datapoint rather than to propose a conceptual rationale for why a discounting parameter might poorly describe an individual's data (Smith & Hantula, 2008). Other researchers have reviewed the impact that exclusion of the nonsystematic datapoints, as determined by the Johnson and Bickel (2008) criteria, have had on their study conclusions, and found very few (White et al., 2015).

In my study, a small but impactful portion of participants' response sets may have been excluded from further analysis if I had followed common practices defining nonsystematic discounting data (Johnson & Bickel, 2008). Therefore, I ultimately retained participant's discounting data irrespective of whether they could be categorised as "systematic" (RQ4.2).

Further to the issue of nonsystematic data, questions about the use of parameters born from model-fitting compared to AUC measures remain important for researchers to investigate and form a suitable position on (e.g., Smith & Hantula, 2008). I used scatterplots and Bland-Altman plots (Bland & Altman, 1986; Bunce, 2008) to investigate the relationship between different measures of discounting (Smith & Hantula, 2008), finding that those derived from model fits are not appropriate for use as primary dependent variables (RQ4.3 & RQ4.4). Specifically, Rachlin's (2006) equation tended to best fit individual response sets but it was not appropriate to compare those equation parameters across participants or conditions. On the other hand, inferential analyses of the raw responses and of the AUC metric derived from them, in this case, both led to the same main conclusion: When the consequence of the scenarios was monetary and comparatively high in severity, participants were more likely to say they intended to intervene than for lower severity scenarios. That said, testing for differences using the AUC metric extended the suite of findings; for example, a significant difference between the consequence types when both were low in severity became apparent. In sum, I determined that AUC was the most appropriate for use as a dependent variable in my studies (RQ4.4).

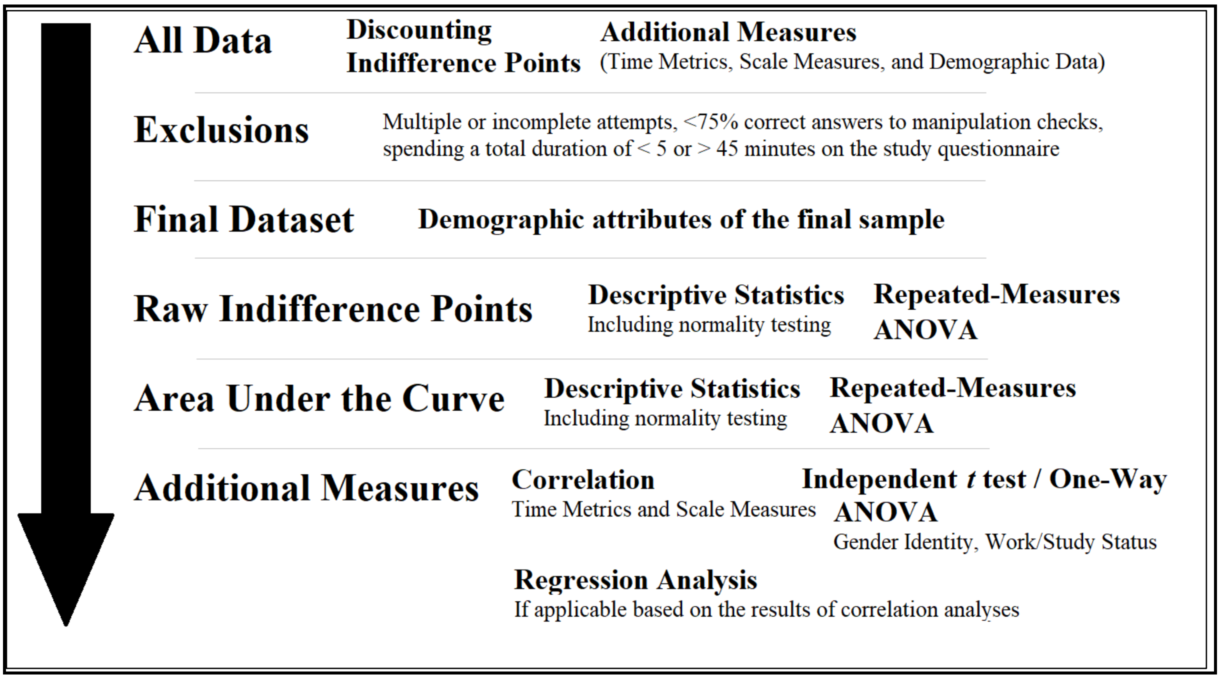
My pilot study (Study 4.1) provides an important proof-of-concept for the application of discounting in regard to online behaviour (RQ4.1 & RQ4.5). Firstly, the likelihood measurement (intention to intervene as a percentage) varied and was sensitive to the different levels of scenario severity (H4.3). Secondly, the intention-to-intervene results were broadly similar for both monetary and social-emotional consequences (H4.1). Emotional or social consequences are less tangible than a commodity (e.g., a sum of money). Thus, it was important to test whether social consequences in a hypothetical online environment were comparable to the use of monetary consequences in discounting procedures (RQ4.5). Finally, most individual and group-level responses were well-described by a hyperboloid (e.g., Rachlin, 2006) discounting function and those that were not tended to have a straightforward conceptual explanation (e.g., always interveners).

A Priori Analysis Plan

As a direct result of the methodological and analytical exploration presented in this chapter, I developed the analysis plan in Figure 6 to implement with each subsequent study, now in an *a priori* fashion.

Figure 6

A Priori Plan for the Analysis of Subsequent Cyberbystander Discounting Data



Summary

I have presented (Chapter 3) and justified (present Chapter 4) a novel application of probability discounting methods to the topic of cyberbystander decision-making (RQ4.1). Importantly, I successfully piloted the use of social-emotional consequences (H4.1) with a discounting procedure using VAS to measure participants' willingness to intervene in online aversive behaviour (RQ4.5). In future studies, I introduce variations of the social-emotional scenarios used in this study (Study 4.1) to better understand the contextual factors that make cyberbystander intervention more or less likely to occur (e.g., Studies 5.1, 5.2, and 6.1). My later studies also serve as additional validation attempts of the application of discounting to cyberbystander behaviour, particularly Studies 5.1 and 5.2 in which I use the same probability scale and similar scenarios to those in Study 4.1. Finally, in subsequent chapters I will compare my results to the framework of the BIM using the discounting methodology established here (Chapter 4).

Chapter 5: Probability Discounting & Cyberbystander Behaviour

Outline

In this chapter, I use the methodology and plan for analysis established earlier (Chapters 3 and 4) with two studies that conceptually extend upon the pilot study. In Study 5.1, I introduced a new independent variable, *responsibility*, to examine how likely a participant was to respond personally (in the *personal* condition) compared to how likely they thought others were to intervene (in the *someone else's responsibility* condition). I retained the pilot study scenarios involving social-emotional consequences (i.e., the ECH and ECL conditions) but manipulated whose responsibility it was to intervene across conditions (personal versus someone else). For Study 5.2, I used the same study design as Study 5.1 but replaced the *severity* variable (i.e., high vs. low severity) with the perceived audience size (*audience*; *small* = 1-2 people; *large* = 50+ people).

By adding the *responsibility* and *audience* variables, I aimed to identify factors that improve the incidence of prosocial online intervention from otherwise passive bystanders. To do this from a behavioural perspective, I reframed Steps 2 and 3 of the Bystander Intervention Model (BIM; Latané & Darley, 1970) and conducted a questionnaire-based discounting study to investigate the utility of these steps for understanding online bystander (e.g., cyberbystanders) behaviour.

In Chapter 4, I showed that discounting methods and the associated analyses can be adequately applied to decision-making when the rewards or consequences are social. My pilot study represents a novel application because, typically, discounting involves the appraisal of commodities rather than social outcomes (see Chapter 4 for further detail). For Studies 5.1 and 5.2, I repeated the social outcome scenarios from my pilot study to further test the hypothesis that these scenarios are suitable for use within discounting procedures. These two predetermined social-consequence scenarios differed according to their level of severity, one moderately high in severity (ECH) and the other moderately low (ECL). Thus, one research question that I examined in Chapter 5 was:

RQ5.1: Are discounting methods useful for those researching bystander behaviour as it functions in the online environment?

I based the scenarios in my series of studies on information collected as part of my exploratory focus groups, which are discussed broadly in Chapter 2. I viewed scenario severity as a continuum ranging from low to high. I reasoned that scenarios at the extremes of the severity continuum would be least ambiguous and, therefore, more easily appraised as requiring intervention. Conversely, it is crucial to investigate scenarios with moderate but potentially harmful levels of severity, as these scenarios are most ambiguous and present bystanders with a more difficult calculus as to whether intervention is needed (Darley & Latané, 1968; Fischer et al., 2011; Macháčková et al., 2015). Therefore, I developed two scenarios closer to the mid-point of the severity continuum to ensure a range of responses and to represent the nuance in social behaviour in online spaces (Orben & Przyblski, 2019; Marciano et al., 2022). I used the two scenarios in the Pilot Study and in Study 5.1, where one scenario was conceptually just below the continuum mid-point (Emotional Consequence - Low Severity; ECL) and the other slightly above (Emotional Consequence - High Severity; ECH). In Study 5.2, I included the severity-scale ranking activity to test the validity of the low- and high-severity scenarios. That is, the ECL and the ECH study conditions, respectively. I hypothesised that:

H5.1: Participants will rank the original (Chapter 4) Emotional Consequence - Low Severity (ECL) scenario lower in severity than the Emotional Consequence - High Severity (ECH) scenario.

Additionally, I added the severity scale to develop more scenarios that cover a broader range of severity levels for future studies. The Study 5.2 severity scale task also allowed me to explore the following research question:

RQ5.2: What severity rankings do participants apply to each of the 20 short scenarios?

The Bystander Intervention Model

Beyond continuing to validate the use of social outcomes in discounting (as in the Pilot Study), I designed Studies 5.1 and 5.2 to investigate the bystander intervention model's second and third steps (BIM; Latané & Darley, 1970). Step 1 of the BIM is beyond the purview of my thesis as it relates to whether potential bystanders notice the inciting incident. My hypothetical scenarios are intentionally short and outline one aversive action rather than being an overview of several behaviours where some are aversive and others are benign. For that reason, all participants "notice" the aversive behaviour because it is the sole behaviour presented in each scenario.

Step 2: Interpret as Requiring Intervention

In a recent systematic review, Mennicke et al. (2022) found a lack of research designed explicitly to examine the second step of the BIM. Latané and Darley (1968) initially described Step 2 as recognising a situation as an emergency. Later work broadened Step 2 to include witnessing behaviour (and not just an emergency) as prompting action from bystanders (Latané & Nida, 1981). Scale measures relating to the BIM steps have most often been developed to investigate the model's final step (i.e., Step 5), taking action (Mennicke et al., 2022). That said, some scales contained a subset of questions related to the interpretation step. For example, a measure of Step 2 from a wider set of items was: "It is evident to me that someone who is being bullied needs help" (Nickerson et al., 2014, p. 284). Because the BIM is linear in its progression, being able to examine Step 2 empirically is at least equal in importance to examining Step 5, in which action is undertaken. I compared my results from Studies 5.1 and 5.2 to Step 2 of the BIM, as deciding to intervene is predicated on believing that intervention is required.

Within the range of responses participants could give, they could indicate that they thought intervention was seldom required. To add to the currently scarce research regarding how often each step of the BIM is fulfilled in cyberbystander scenarios, I explored the effect of setting an intending-to-intervene threshold of 50% or more to fulfil Step 2, as established in Chapter 3. In other words, participants who showed an overall moderate-to-high intention to intervene were indicating that the scenario required intervention. Therefore, they fulfilled Step 2 of the BIM.

RQ5.3: What proportion of participant responses fulfil Step 2 of the BIM?

Furthermore, if Step 2 of the BIM were a critical part of participants' decision-making, I expected behavioural intentions to differ across the severity manipulation. Research about the bystander effect indicates that the more severe an incident is, the more likely bystanders will intervene (e.g., Fischer et al., 2011; Phares & Wilson, 1972). In a meta-analysis, Janz and Becker (1984) found that a bystander's perception of scenario severity was significantly correlated with behaviour in 65% of their reviewed studies. That said, sometimes these correlations were negative, suggesting that severity sometimes led to more precautionary behaviour (e.g., standing by) rather than proactive behaviour (e.g., intervening to defend a target; Janz & Becker, 1984). In a more recent meta-analysis, Sheeran et al. (2014) found a medium-to-large effect of risk appraisal on bystanders' resultant behaviour. These findings support research regarding the severity of a scenario witnessed online and on identifying the environmental factors that augment cyberbystanders' perception of severity. To begin exploring the impact of scenario severity, I used the Study 5.1 dataset to test the following hypothesis:

H5.2: In Study 5.1, the intention to intervene will be most likely when the outcomes in the scenario are more severe.

Note that H5.2 is contingent on having first found support for H5.1, that participants generally view the low (ECL) and high severity (ECH) conditions as such.

Another aspect related to Step 2 of the BIM is the ambiguity often present in the aggressors' intentions. That is, whether the aggressor intended to cause harm (Olewus, 1997).

Whether an aggressor intended to cause harm sometimes varies in line with how clear the severity level of a situation is to bystanders. Consider an instance where someone's ex-partner has carried out a vicious, personalised attack. In this case, there is likely to be less ambiguity both in terms of the severity level (i.e., highly severe) and whether the aggressor intended to cause harm. However, there are also cases where intentionality and severity are not parallel. For example, the aggressor from my previous example could be anonymous rather than being identified as the target's ex-partner. In this case, the severity level is still clear, but the intentionality less so. To test the impact of ambiguity as it related to whether harm inflicted was intentional, I hypothesised that:

H5.3: In both studies, the likelihood of intervening will increase the more certain it is that the aggressor intended to cause harm.

Step 3: Accepting Responsibility for Intervention

As Latané and Darley (1968) put it, Step 3 of the BIM is "to assume the responsibility to take action" (p. 220), which, in my studies, means to intervene. In behaviourist terms, accepting responsibility is most evident by carrying out an intention (e.g., intervening). In my studies and several others (e.g., Mennicke et al., 2022), a measure of participant intention is a necessary proxy for the actual behaviour of intervening. There is mixed evidence regarding how closely a measure of behavioural intention relates to actual behaviour (e.g., Bhattacherjee & Sanford, 2009; Sheeran, 2002; Sheeran et al., 2014). Nevertheless, it is possible that discounting methods are robust to the impact of using hypothetical measures (i.e., intention to intervene) and scenarios, rather than real ones, on study outcomes (Johnson & Bickel, 2002). Moreover, researchers have found support for the equivalence of real and hypothetical outcomes in probability discounting (Matusiewicz et al., 2013) and social discounting (Locey et al., 2011), which are both relevant to my work.

Utilising hypothetical scenarios, I manipulated whether participants answered on their behalf (the personal condition) or on behalf of another bystander (someone else's responsibility condition). I examined the difference between those who believed intervention was required (Step 2) and those who also assumed personal responsibility for intervening (Step 3; Latané & Darley, 1968). For example, participants who gave high likelihood-to-intervene responses in the someone else's responsibility conditions, as compared to the personal ones, fulfil Step 2 of the model but do not accept responsibility. As with Step 2, I considered participants who reported a moderate-to-strong overall likelihood (i.e., 50% or greater) of intervening in the personal conditions as fulfilling Step 3. More generally, the question I assessed was:

RQ5.4: What proportion of participants fulfil Step 3 of the BIM?

For all scenarios in Study 5.1, I stated that other people also witnessed the aversive behaviour. Therefore, there were always other people (i.e., an audience) whom a participant could presume would intervene even if they themselves were not going to. One explanation when there is a lack of bystander intervention is that the presence of others leads one to reappraise the event as not requiring intervention. In other words, bystanders reason that if an intervention were needed, someone else would have already done so (Fischer et al., 2011). Darley and Latané (1968) described this phenomenon as a *diffusion of responsibility*. That is, an individual within a group tends to look to others as part of their decision-making process regarding whether to intervene (e.g., Rabb et al., 2022). If others do not react to the situation, group members who may have otherwise intervened infer that their peers believe intervention is unnecessary (Butler et al., 2022; Fischer et al., 2011). Researchers have found that the presence of others can lessen the chance of bystander and cyberbystander intervention (e.g., Abbate, 2022; Campos-Mercade, 2021; Hussain et al., 2019).

Previous researchers suggest that at a certain size, group size has less impact on behaviour (Allison & Bussey, 2016). Just as differences between longer delays (e.g., 365 and 370 days) are less discriminable than the same difference at shorter delays (e.g., 1 compared to 6 days) in delay discounting, large numbers of other bystanders seem to be interpreted in a similar way. Obermaier et al. (2016) found that their participants rated 24 bystanders as a moderately sized audience, while audiences of 224 and 5025 were rated large and not considered meaningfully different despite a marked disparity in the number of people present. For this reason, I included a two-level manipulation of audience size in Study 5.2, one where the audience size was only two other people (small) and one with an audience of 50 or more other people (large).

Regarding my audience size manipulation in Study 5.2, I posited that:

H5.4: The intention to intervene in Study 5.2 will be less likely when the perceived audience size is large.

In addition to severity level, whether harm is intended, and audience size, prosocial intervention may be viewed in two other ways that relate to whose responsibility it is to intervene. One is that an active bystander risks the aggressor retaliating against them in response to their interference (Abbate et al., 2022). That is, they could become a new target for aversive behaviour (i.e., positive punisher), which may result in people being less likely to intervene. It could also be that a bystander chooses to avoid intervening entirely, which constitutes a negative reinforcer as they are removing their chance of being harmed (Dymond et al., 2014; Schlund et al., 2021). A substantial body of research suggests that in the face of threatening behaviour, a sizeable proportion of people exhibit avoidance behaviours (Dymond & Roche, 2009; Schlund et al., 2021; Sidman, 1955) but are less likely to factor in the impact of negative reinforcers on the behaviour of others (Allport, 1924; Jones & Nisbett, 1987; Weatherly et al., 1999).

Another way to view prosocial intervention is that it may be appreciated as a show of support, regardless of whether the target needed assistance (i.e., if no harm was done nor intended). For the bystander, intervening in the defence of others may help them garner social reinforcement and support their sense of being a good or superior person (Codol, 1975). Put another way, if people believe themselves to be better than the average person, intervening to help could be the most congruent behavioural response and would be the option with the most positive reinforcement available (Alicke et al., 1995; Brown, 2012); therefore, participants would rate themselves as more likely to intervene than others. Subsequently, an observable show of support - regardless of whether it was required - could function to establish such prosocial bystander behaviour as normative within online spaces (e.g., Crone et al., 2022; Rabb et al., 2022). The social context greatly influences behaviour; therefore, the more visible prosocial intervention is, the more opportunities there will be for others to see, learn from, and model such behaviour (Baum, 2017; Wolter & Napper, 2023). In sum, there is evidence in both directions. Therefore, Hypothesis 5.5 is bidirectional:

H5.5: In both studies, the intention to intervene will differ as a function of whether the responsibility for intervening is that of the participant or of someone else.

Additional Measures

In the results from my pilot study (Chapter 4), participant intention to intervene was, in some cases, correlated with additional measures I included as possible covariates. To further explore those, I include a research question for later discussion:

RQ5.5: In what ways (if any) is the intention to intervene related to other variables (time online, time on social media, proportion of online time that is on social media, smartphone addiction, perceived socio-economic status, and/or self-esteem)?

In addition to the possible covariate variables included in the Pilot Study, I aimed to investigate the relationship between smartphone addiction and cyberbystander behaviour in subsequent studies. I reviewed a number of scales on the general topic of smartphone use (Harris et al., 2020). In particular, the shortened Smartphone Addiction Scale (SAS; Kwon et al., 2013) has been reliable (Demirci et al., 2014; Marciano & Camerini, 2022; Nikolić et al., 2021) and valid in several contexts (e.g., Alhassan et al., 2018; Demirci et al., 2015; Tateno et al., 2019). Also, the shortened version has been shown to predict the long version of the scale and have good concurrent validity (SAS-SV; Kwon et al., 2013). Therefore, I used the shortened scale (hereafter only referred to as SAS).

Current Studies

The current chapter comprises two studies (Studies 5.1 and 5.2) that have several commonalities. In both, I used a probability discounting procedure to test whether participants view their likelihood of intervening in an aversive online situation as equal to that of others (H5.5) and also whether the likelihood of intervening decreases as a function of intention-related ambiguity (H5.3). In Study 5.2, I added an extra set of questions designed to validate my manipulations of scenario severity (H5.1) and its impact on cyberbystander intentions in the Pilot and in Study 5.1 (H5.2). The extra scenario-severity questions were also useful for examining a wider range of scenarios, including how participants viewed their severity

(RQ5.2). Also, in Study 5.2, I included a manipulation of the perceived audience size to assess whether participant intention to intervene decreased in the presence of increasing numbers of others (H5.3). In addition to the five testable hypotheses, I compared the results of Studies 5.1 and 5.2 to the second and third steps of the BIM (RQ5.3 and RQ5.4, respectively). Step 2 relates to whether cyberbystanders perceive that a situation requires intervention. Then, Step 3 is whether cyberbystanders personally accept the responsibility for intervening (Latané & Darley, 1970). Finally, the overarching question for my thesis was whether discounting methods are useful for understanding cyberbystander behaviour (RQ5.1) and, if so, how these can best be applied.

Methods

Recruitment

For Study 5.1, I used two different recruitment methods. First, I recruited participants from a cohort of 1st-year psychology students at an Aōtearoa New Zealand university. Secondly, I recruited participants from the online research recruitment platform Prolific (www.prolific.co). For Study 5.2, I recruited all participants using the Prolific platform.

On Prolific, I applied two broad conditions for participation: (1) that participants be between 16 and 65 years, and (2) that they report using the internet at least twice weekly. I verified that University student participants met this criterion using their self-reported time online data.

All participants received a small reward for taking part. Participants recruited from Prolific received payment at a hourly rate (USD 6.50) calculated pro rata and per the platform's guidelines at that time (www.prolific.co; accessed 2020). University student participants received course credit in proportion to the time required to complete my study, a process facilitated by University staff. Recruitment and data collection for Study 5.1 occurred from May to June 2021 and throughout August 2021 for Study 5.2.

Sample

The questionnaire for Study 5.1 was begun by 196 individuals. Of those, 15 of the students (i.e., Uni) began but did not complete my questionnaire. A further two Prolific participants and three students took less than 5 minutes to complete the questionnaire, a duration that I had pre-emptively set as the minimum time in which my study could be satisfactorily completed. Two of the Prolific participants took longer than 45 minutes to complete my questionnaire, which exceeded the upper threshold I had set for an adequate response. A further eight participants answered fewer than 75% of the manipulation-check questions correctly. Therefore, the final sample size of Study 5.1 was 163 participants, and those excluded did not differ significantly from those retained based on either their demographic characteristics or psychometric scale scores.

The sample I used in Study 5.1 included psychology students (n = 69) and those recruited on the Prolific research platform (n = 94). Before I combined these groups for subsequent analyses, I tested whether the student and Prolific participants were sufficiently similar across all the scale variables (i.e., the discounting AUC values for each condition, duration, the extent of nonsystematic discounting, TO, ToSM, PoT, PSES, RSE, SAS, and age). The groups only differed regarding age; participants recruited on Prolific tended to be older than those recruited from the university population, $F(1, 162) = 8.20$ $p = .01$. On average, participants in Study 5.1 were in their early twenties (Range = 17-52 years, $M_{Uni} = 22$, $SD_{Uni} = 8$; $M_{Prolific} = 25$, $SD_{Prolific} = 8$). Overall, age was moderately upwardly skewed, resulting in a cluster of participants who were younger or exactly the age of their respective group averages (*Uni* = 74%, *Mdn* = 19; *Prolific* = 66%, *Mdn* = 23). In addition, most university student participants in the sample selected “female” as their gender identity (82.19%), whereas the subset of Prolific participants was more evenly distributed across the categories of gender identity. Therefore, I tested whether gender was a confounding variable at the outset and found no statistically significant differences as a function of participant gender identity. Overall, though, the University and the Prolific participants were comparable; thus, I combined these groups into one sample.

The questionnaire for Study 5.2 was begun by 135 individuals. One person completed the questionnaire twice, so I retained only their first attempt in the final sample. Another individual began but did not complete my questionnaire. A further 17 people answered fewer than 75% of the manipulation-check questions correctly. The incorrect answers given were relatively evenly spread between the questions designed to check the audience-size manipulation and those for checking the responsibility-for-intervening manipulation. One further participant took more than 45 minutes to complete the questionnaire. Therefore, the final sample for Study 5.2 was 115 participants.

The average age of participants in Study 5.2 was 25 years old (*SD* = 8), although this was skewed positively, and the median age was lower (*Mdn* = 22). In this sample, a majority of participants self-reported their gender identity as female (*n* = 80, 69.6%). A further 27.0% identified as male (*n* = 31) and 3.5% as non-binary (*n* = 4).

Apparatus

Participants recruited from the Prolific platform (www.prolific.co) completed the Qualtrics questionnaire after reading a brief description of the study. University-based participants arrived at the same questionnaire through the SONA research platform that the School of Psychology used for administering undergraduate course credit. The Qualtrics questionnaire could be completed on any internet-enabled device at any time that suited the participant.

Procedure

The procedures of Studies 5.1 and 5.2 have more similarities than differences. Unless otherwise noted, the same procedure applied to Studies 5.1 and 5.2.

Discounting

All participants first completed 24 discounting trials, six in each of four conditions. Each condition included a different description of an aversive scenario of the type one could witness on SNS. I kept the text of each scenario as similar as possible across conditions in regard to length and language used, and gender neutrality. Two conditions measured how the participants themselves would behave, while the other two measured how the participants believed that other people would behave (i.e., responsibility for intervening). To ensure this responsibility manipulation was salient, the personal scenarios were preceded by the following note (emphasis included): “For the next scenario, please answer the six (6) questions according to how ***you*** would behave”.

Similarly, the scenarios in which someone else was responsible for intervening (i.e., someone else conditions) were preceded by:

> For the next scenario, please answer the six (6) questions according to how **a generic other person** would behave, e.g., how **an acquaintance** of yours might respond. For the purpose of this section, imagine that this acquaintance is **named “Zee”**.

For each of the six trials within each condition, I asked participants to imagine that they had varying levels of certainty as to whether the aggressor was intending to cause harm. My certainty-of-harm intentionality measurement was probabilistic and presented to participants as a percentage (%). For example, “You are about 30% sure”. Every trial in the discounting section of my questionnaire concluded with the same general prompt question: “How likely are you to intervene?” Participants responded to each prompt by moving an indicator on a visual analogue scale (VAS) that ranged from 0%, labelled as “Unlikely to reply,” through to 100%, “Will definitely reply” (e.g., Brumfitt & Sheeran, 1999; Kaplan et al., 2014).

In Study 5.1, I manipulated the scenario severity in addition to whose responsibility it was to intervene. The high-severity and personal-responsibility scenario was:

> Imagine that you are online when you come across a nasty rumour about a close friend of yours. A recent ex-partner of your friend has accused them of treating the ex badly and being wholly responsible for the breakdown of the relationship. The ex-partner has posted publicly and so the accusations are easily viewable to your friend, as well as all their friends and family.

For the corresponding high-severity scenario, in which the responsibility to intervene was someone else’s, I asked participants to read the following and imagine what Zee might do:

> Imagine that Zee is online when they come across a nasty rumour about a close friend of theirs. A recent ex-partner of Zee's friend has accused them of treating the ex badly and being wholly responsible for the breakdown of the relationship. The ex-partner has posted publicly and so the accusations are easily viewable to Zee's friend, as well as all their friends and family.

The low-severity and personal-responsibility scenario in Study 5.1 was:

> Imagine that you see a social media post where it appears that another user has an issue with a close friend of yours. Your friend is being accused of being rude because they completely ignored the person posting. The poster of the message apparently had said "Hello" to your friend, trying to be friendly, at a recent party. Knowing your friend, you are sure that it would have been accidental and think your friend would be embarrassed to know that they had offended someone.

The low-severity scenario that required participants to imagine how someone else may respond (i.e., Zee) was:

> Imagine that Zee sees a social media post where it appears that another user has an issue with a close friend of theirs. Zee's friend is being accused of being rude because they completely ignored the person posting. The poster of the message apparently had said "Hello" to Zee's friend, trying to be friendly, at a recent party. Knowing their friend, Zee is sure that it would have been accidental and thinks their friend would be embarrassed to know that they had offended someone.

I designed Study 5.2 to replicate the manipulation of whose responsibility it was to intervene, the participant's own, or someone else's. In addition to the responsibility, I tested the effect of the perceived audience size in Study 5.2. The scenario with the large audience and that required participants to consider whether they would intervene was:

> Imagine that you are online when you come across a nasty rumour about a close friend of yours. A recent ex-partner of theirs has posted on social media, accusing your friend of being wholly responsible for the breakdown of the relationship. They have posted publicly to a space where your friend is all-but guaranteed to see the post, along with upwards of 50 other onlookers.

The corresponding Study 5.2 scenario, with a small-sized audience, was:

> Imagine that you are online when you come across a nasty rumour about a close friend of yours. A recent ex-partner of theirs has posted on social media, accusing your friend of being wholly responsible for the breakdown of the relationship. They have posted publicly to a space where your friend is all-but guaranteed to see the post but it is unlikely that others will, perhaps just you and one other onlooker.

As in Study 5.1, these conditions were also presented to participants in a way that asked them to consider how someone else might respond to the scenarios. The order of conditions was randomised across the sample to control for possible order effects. I present the full, in-order sequence of scenarios and questions for both studies (5.1 and 5.2) in Appendices D and E for reference.

Additional Measures

I included measures to investigate potential correlates of cyberbystander behaviour and to validate existing study resources. As in the Pilot Study (Chapter 4), Studies 5.1 and 5.2 included questions regarding participants' time spent online and on social media, the 10-item Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965), the single-item MacArthur Subjective Social Status scale (PSES; Adler et al., 2000), and basic demographic items (i.e., gender identity, employment or study status, and year of birth). To Study 5.2, I added a 10-item scale intended to measure smartphone addiction (SAS; Kwon et al., 2013) and a 10-item severity-ranking activity that I developed (further information on additional measures presented in Appendix C).

I included both the short-form Smartphone Addiction Scale (SAS-SV) and participant self-reports of their own time on social media (as a subset of overall time online) as a partial replication of previous work (Ellis et al., 2019) and to assess whether these metrics were congruent.

The severity-scale ranking activity that I included in Study 5.2 had 20 sentence-length scenarios of varying severity. I shortened the low- and high-severity scenarios from my Pilot Study (the ECH and ECL scenarios) and the two studies presented in this chapter (Study 5.1 and 5.2) to test whether participants rated these as low and high in severity, respectively. The short-form of the ECH scenario was: "An ex-partner is spreading untrue and hurtful information about Y online," and I coded this as *ExRumour*. The shortened version of the ECL scenario was: "X posted online to accuse Y of being rude to them at a recent party, even though Y was not present," which I coded *PartyRudeness*. A further ten of the severity statements were closely adapted from the work of Hellström and Lundberg (2020). The remaining eight statements I adapted from the results of my focus groups.

The severity-ranking activity was optional in Study 5.2, although most did complete it (97.74%). Each participant was presented with a total of 10 short statements that they could drag and reorder to reflect how severe they considered the scenario to be. The *ExRumour* and *PartyRudeness* statements were always included. The other eight statements were randomly selected from those remaining. The initial order of the statements was randomised. Moreover, the extra items were to aid in developing a more nuanced scale version (rather than dichotomous) for subsequent studies.

In summary, all participants first completed the discounting section, followed by the self-reported time-spent section (i.e., TO and ToSM). Those who participated in Study 5.1 then completed a self-esteem scale (RSE; Rosenberg, 1965), a perceived socioeconomic status question (PSES; Adler et al., 2000), and basic demographic information. Those who participated in Study 5.2 completed the same measures, with the addition of the smartphone addiction scale (SAS; Kwon et al., 2013) and the severity scale ranking activity. In both studies, the order that the additional measures (RSE, PSES, SAS, and the severity scale) were completed was randomised. Within each question set, the individual items were also randomised, except for PSES, which contains a single item.

Analysis

As established in a prior section (Chapter 3 and 4), my plan for the analysis of this dataset was set out in advance and is summarised in Figure 6.

Findings & Discussion

Study 5.1: Responsibility X Severity

Discounting Methods for Cyberbystander Research (RQ1 & H3)

Table 8 has descriptive statistics of the raw indifference point data for all conditions and certainty values. When the certainty that harm was intentional was stated to be 50% or lower, participants tended to comparatively over-weight the intention to intervene. Put another way, intention to intervene was more likely than the corresponding certainty value would, if matched, predict (e.g., at 50% certainty, intention to intervene was close to 60% for the O.ECH condition). At most certainty points (except for 5%), intention to intervene was highest within the O.ECH condition. That is, participants indicated that intervention was most likely to occur when it was someone else's responsibility and when the scenario was more severe.

Table 8

Descriptive Statistics at each Certainty Value and for each Study Condition in Study 5.1

	95%				90%				50%				30%				10%				5%			
	P.ECH	O.ECH	P.ECL	O.ECL	P.ECH	O.ECH	P.ECL	O.ECL	P.ECH	O.ECH	P.ECL	O.ECL	P.ECH	O.ECH	P.ECL	O.ECL	P.ECH	O.ECH	P.ECL	O.ECL	P.ECH	O.ECH	P.ECL	O.ECL
M	76.85	85.58	77.79	83.70	74.72	83.36	75.23	80.67	53.75	59.02	52.66	54.81	41.64	44.99	40.80	40.65	32.15	33.42	30.41	31.15	29.75	28.42	26.67	26.07
SD	31.90	22.60	29.15	24.04	31.34	18.89	28.52	23.02	28.52	18.06	26.38	20.27	28.46	24.06	28.96	21.74	31.07	28.18	29.61	26.24	32.84	29.81	30.60	28.94
SEM	2.50	1.77	2.28	1.88	2.45	1.48	2.23	1.80	2.23	1.41	2.07	1.59	2.23	1.88	2.27	1.70	2.43	2.21	2.32	2.06	2.57	2.33	2.40	2.27
Mdn	92	95	90	95	89	90	84	89	60	60	58	55	39	40	40	39	20	21	20	22	18	17	15	13
Skew	-1.33	-2.37	1.20	2.61	-1.31	-1.90	1.29	2.72	-0.41	0.08	-0.50	0.19	0.35	0.45	-0.83	-0.63	0.83	0.88	-0.33	-0.66	0.91	0.96	-0.04	-0.26
Kurt	0.33	5.38	-1.50	-1.85	0.39	3.95	-1.48	-1.79	-0.75	0.22	-0.30	-0.28	-0.80	-0.73	0.33	0.31	-0.54	-0.41	0.92	0.71	-0.53	-0.30	1.09	1.04

Note. 95%, 90%, 50%, 30%, 10%, and 5% are certainty values, *M* = mean indifference point, *SD* = standard deviation, *SEM* = standard error of the mean, *Mdn* = median indifference point, *Skew* = skewness, *Kurt* = kurtosis of the distribution of the indifference points (*N* = 163). P.ECH = personal responsibility, high severity; O.ECH = someone else’s responsibility, high severity; P.ECH = personal responsibility, low severity; O.ECL = someone else’s responsibility, low severity.

Across all conditions, most participants intended to intervene at high levels of certainty (i.e., 95%) and became less likely to intervene as the certainty decreased (Table 8). In other words, most participants discounted the benefit of intervening as a function of decreasing certainty. At the group level, the rate of discounting was best described by Rachlin's (2006) hyperboloid function. In every condition, at least 80% of the individual response sets were best-fit by Rachlin's (2006) equation (P.ECH = 80.37%; O.ECH = 85.28%; P.ECL = 85.28%; O.ECL = 90.80%). Within those response sets that were well-described by Rachlin's (2006) equation, the sensitivity parameter was best set at 1.00 for approximately a quarter of participants (P.ECH = 21.47%; O.ECH = 28.22%; P.ECL = 22.70%; O.ECL = 26.99%); thus, Rachlin's (2006) equation is effectively reduced to a single parameter that mirrors Mazur's (1987) equation for those participants.

In sum, I found support for Hypothesis 5.3: Participant response sets showed that the utility of intervening was discounted as a function of decreasing certainty (and increasing ambiguity) in how to interpret the scenario. Moreover, because participant data were satisfactorily fit by common discounting functions (e.g., Mazur, 1987; Rachlin, 2006), my results suggest that discounting methods can be usefully applied to cyberbystander research (RQ5.1).

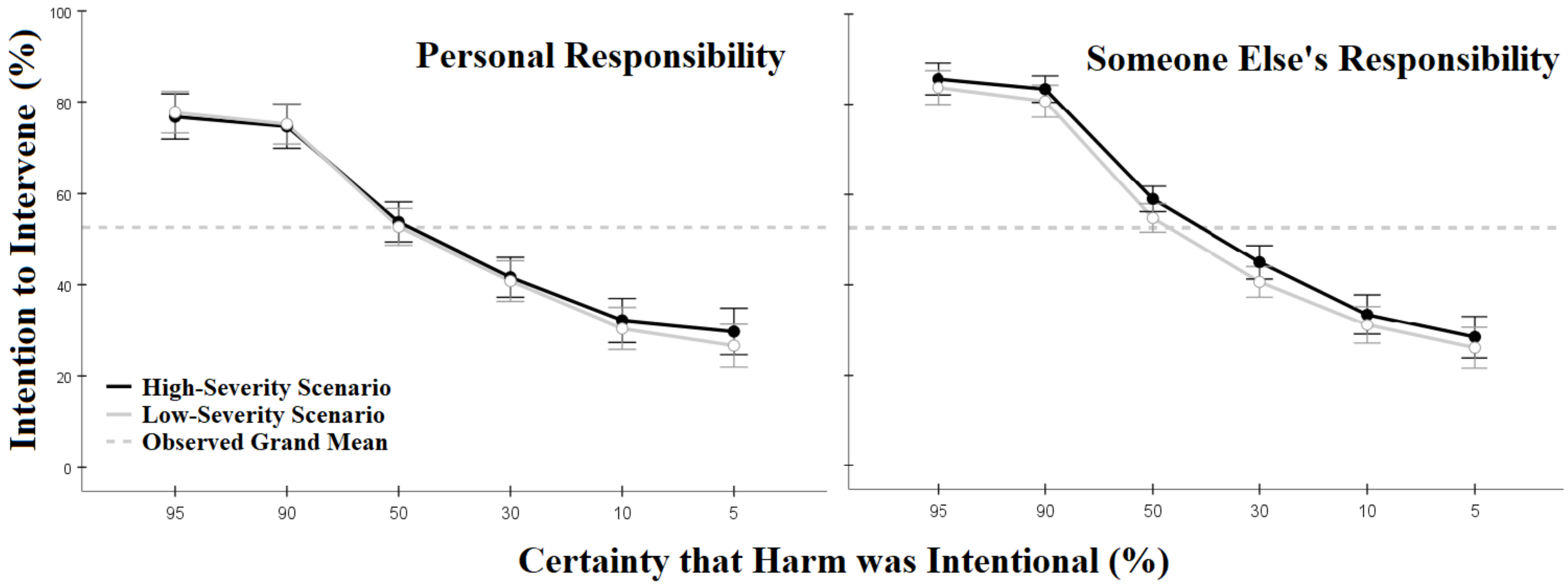
Scenario Severity (H2) & Responsibility for Intervention (H5)

To test for effects of the main manipulations (i.e., severity, H5.2; responsibility, H5.5) in Study 5.1, I first conducted a repeated-measures ANOVA using the raw indifference points from participants as the dependent measurement. Although some of the test assumptions were not wholly met, such as the distribution of data approximating normal, $W(5) = .52$, $p < .001$, the ANOVA identified a significant main effect of condition after Greenhouse-Geisser corrections, $F(3, 486) = 4.46$, $p = .01$, $R^2 = .03$. Post-hoc pairwise comparisons revealed that the only differences between conditions were between the O.ECH scenario and others. The

O.ECH and O.ECL scenarios, which differed in the level of severity, generated significantly different intention-to-intervene responses as shown in the right panel of Figure 7, M_{Diff} = 2.96, 95% CI [.68, 9.08], p = .02. Similarly, responses to the O.ECH and P.ECL conditions significantly differed, M_{Diff}= 5.21, 95% CI [.68, 9.73], p = .02. There was no significant difference between responses to the two high-severity scenarios (O.ECH and P.ECH), 95% CI [-0.44, 9.08], p = .10. There was also a significant main effect of certainty, $F(5, 810)$ = 446.23, $p < .001$, $R^2 = .73$, and an interaction effect between condition and certainty, $F(15, 2430) = 3.25$, $p = .002$, $R^2 = .02$.

Figure 7

Intention to Intervene as a Function of Certainty in Study 5.1



Note. 95% confidence intervals are indicated by the error bars. Grey lines with open circles = the scenarios low in severity; the black lines with closed circles = the scenarios comparatively higher in severity; the dotted grey line indicates the overall average intention to intervene.

AUC was a more appropriate metric for describing cyberbystander-related discounting (see Chapter 4 for a review). Major findings, such as that intervention was most likely in the O.ECH condition, were the same regardless of whether I used parameters from models (i.e., *h* and *s*) or AUC metrics. Furthermore, the AUC metric is more suitable for use in parametric statistical tests, as shown by the descriptive statistics in Table 9. For example, both the skewness and the kurtosis of AUC in each study condition support that AUC is normally distributed, which is in direct contrast to the unusual and skewed values present within the model-based parameter data.

Table 9

Area Under the Curve Descriptive Measuring for Each Condition in Study 5.1

	P.ECH	O.ECH	P.ECL	O.ECL
M	0.532	0.581	0.527	0.547
SD	0.257	0.169	0.242	0.173
Skew	-0.341	0.286	-0.293	-0.497
Kurt	-0.405	0.762	-0.113	0.906
Mdn	0.552	0.547	0.531	0.538

Note. P.ECH = personal responsibility, high severity; O.ECH = someone else's responsibility, high severity; P.ECH = personal responsibility, low severity; O.ECL = someone else's responsibility, low severity.

A 2-x-2 repeated-measures ANOVA using the personal-versus-someone-else conditions (i.e., responsibility) and the high-versus-low conditions (i.e., severity) identified a main effect of responsibility, $F(1, 162) = 5.16$, $p = .02$, and severity, $F(1, 162) = 4.91$, $p = .03$, but no interaction between responsibility and severity, $F(1, 162) = 3.23$, $p = .07$. On average, the difference between responses in the high-and low-severity scenarios was 2.0%, CI 95% [0.2, 3.7], $F(1, 162) = 4.91$, $p = .03$, $R^2 = .03$, and AUC was greatest when the scenario was high in severity (M_{High} = 55.7%, 95% CI_{High} [52.8, 58.5]; M_{Low} = 53.7%, 95% CI_{Low} [50.9, 56.5]). When participants were asked to respond as to how likely they - rather than someone else - would be to intervene, there was a small but significant difference, *MDiff* = 3.5%, 95% CI [0.5%, 6.6%], $F(1, 162) = 5.16$, $p = .02$, $R^2 = .03$, such that they rated their own intention to intervene as lower (*M* = 52.9%, 95% CI [49.3, 56.6]) than when the responsibility was someone else's (*M* = 56.4%, 95% CI [54.0, 58.8]). These results support Hypothesis 5.2, that intention to intervene in aversive situations online is greatest when the severity is high, and Hypothesis 5.5, that participants believe someone else to be more likely to intervene.

Fulfilment of the BIM Steps

Further to the group-level results, the rate at which individual participants rated their own and someone else's intentions to intervene was of interest. For Table 10 I first calculated each participant's average likelihood of intervention for each condition (i.e., the sum of the six indifference points divided by six). I then separated the average likelihood values into two categories: those who were most likely to say that intervention would occur (50% or more) and those comparatively less likely to (less than 50%). In all conditions, at least a slight majority of participants indicated that they thought intervention was more likely to occur than not (i.e., that there was a 50% or more chance of intervention); thus, appear to fulfil Step 2 of the BIM and provide an estimate regarding RQ3. Step 2 is that bystanders recognise a situation as requiring intervention.

Table 10

Participants in Study 5.1 who Indicated that They, or Someone Else, Would Intervene

	P.ECH	O.ECH	P.ECL	O.ECL
Less than 50%	[a]74	65	81	74
	[b]45%	40%	50%	45%
50% and greater	89	98	82	89
	55%	60%	50%	55%

Note. Frequency count of the participants who had an average indifference point of 50% or greater (lower panel) or an indifference point of less than 50% (upper panel). [b]The frequency count for each category presented as a percentage of the total sample. P.ECH = personal responsibility to intervene, high severity scenario; O.ECH = someone else's responsibility to intervene, high severity scenario; P.ECL = personal responsibility to intervene, low severity scenario; O.ECL = someone else's responsibility to intervene, low severity scenario.

In regard to Step 3 of the BIM and RQ5.4, fewer Study 5.1 participants rated intervention as likely (i.e., 50% or more of the time) when it was their personal responsibility to do so (Total Personal = 52%; P.ECH = 55%; P.ECL = 50%), compared to when they believed it was someone else's responsibility (Total Someone Else's = 57%; O.ECH = 60%; O.ECL = 55%). Step 3 relates to whether a bystander takes personal responsibility for intervening. In my results, 52% of participants, across both personal conditions, were willing to take responsibility and intended to intervene in aversive behaviour personally. An additional 5% of participants responded that they thought someone else would intervene (57%) when they themselves would not. Put another way, there is less evidence that participants would fulfil Step 3 (i.e., take personal responsibility for intervening) and comparatively more evidence that participants would fulfil Step 2 (i.e., recognise the situation as requiring intervention). Furthermore, the average intention to intervene when the responsibility was the participants was significantly lower (46.08%, 95% CI [45.08, 46.96]) compared to when the responsibility was someone else's (66.43%, 95% CI [65.55, 67.31]), $t(325) = -2.78$, $p = .01$, $d = -.15$. These results support the claim that individuals taking personal responsibility for carrying out cyberbystander intervention is an important component of the decision-making process (Latané & Darley, 1968), although studies designed explicitly to test this claim are needed.

Using a chi-square test of independence, I found no statistically significant relationship between the study conditions and whether a participant's average indifference point was below 50%, or 50% and above, $X^2(3) = 3.20$, $p = .36$, $\Phi = .07$. That said, the VAS scale in Studies 5.1 and 5.2 ranged from "Unlikely to reply" through to "Will definitely reply" and so identifying participants who would not intervene (i.e., would not have replied) was not possible. I intended RQ5.3 and RQ5.4 to be exploratory in testing whether cyberbystanders achieved Steps 2 and 3 of the BIM. Using descriptive analyses, it may be cautiously theorised that approximately 55% of Study 5.1 participants fulfilled Step 2, in that they interpreted the scenario as requiring intervention and therefore intended to provide that intervention (or believed that someone else would). More participants rated intervention as likely when the responsibility to intervene rested with someone else (57%) than when they were responsible for intervening (52%); thus, I preliminarily estimate that 52% of my sample fulfilled Step 3. In future research, this finding should be tested using an extended VAS scale that would more clearly measure a desire not to intervene. For example, a VAS with three labels where zero is "Would definitely not reply", the mid-point (i.e., 50) is a neutral point for the undecided participant, and 100 remains labelled "Will definitely reply".

To summarise, I made a cautious initial estimate of the proportion of participants who fulfilled Step 2 (RQ5.3 = 55%) and Step 3 (RQ5.4 = 52%) of the BIM, when applied to cyberbystander behaviour. More important, though, the process by which I calculated the estimates appears to be a useful starting point for testing steps of the BIM that have yet to be comprehensively examined (Mennicke et al., 2022).

Cyberbystander Intervention Covariates (RQ5)

I included several additional variables in Study 5.1 that may have covaried alongside the propensity to discount cyberbystander intervention. Of the possible covariates, the variable that was the most highly correlated with the other covariates was the self-reported time on social media (ToSM) measurement. Of all the possible relationships between variables, displayed in Table 11 for reference, all but PSES and the O.ECH condition were significantly correlated with ToSM. Other than general time spent online (TO), ToSM was most strongly related to the construct of smartphone addiction (r_{SAS} = .41, p <.001), then, to a lesser extent, age (r = -.27, p <.001), with a small-sized relationship to self-esteem (r_{RSE} = -.17, p = .03). For so-called smartphone addiction, as time on social media increased, so did scores on the SAS, which indicates more intensive smartphone addiction. I also found a trend suggesting that the older a participant was, the less time they spent on social media. The relationship between RSE and ToSM was also negatively correlated, indicating that self-esteem tended to decrease as social networking time increased. In addition, ToSM was positively associated with the AUC values from three of the four study conditions ($r_{P.ECH}$ = .23, p = .003; $r_{P.ECL}$ = .29, p = <.001; $r_{O.ECL}$ = .16, p = .04). In other words, for most conditions, it was the participants who spent more time on social media who were also most likely to intervene.

Table 11

Correlations Between Additional Measures and the Intention to Intervene AUC Values from Study 5.1

	TO	ToSM	PSES	RSE	SAS-SV	Age	P.ECH	O.ECH	P.ECL	O.ECL
TO							0.038	-0.041	0.061	-0.054
							0.633	0.601	0.441	0.490
ToSM	**0.298**						**0.231**	0.134	**0.288**	**0.163**
	<.001						**0.003**	0.087	**<.001**	**0.037**
PSES	-0.048	-0.086					0.104	-0.069	0.109	0.080
	0.541	0.274					0.187	0.382	0.166	0.311
RSE	-0.102	**-0.171**	0.144				-0.031	-0.054	-0.040	-0.013
	0.194	**0.029**	0.066				0.693	0.490	0.608	0.867
SAS-SV	0.096	**0.405**	0.042	**-0.204**			0.154	**0.192**	0.152	0.200
	0.223	**<.001**	0.598	**0.009**			0.050	**0.014**	0.052	0.010
Age	0.058	**-0.273**	-0.089	0.154	-0.213		-0.101	**-0.155**	-0.089	-0.012
	0.460	**<.001**	0.261	0.050	0.006		0.202	**0.048**	0.258	0.875
PoT	**-0.312**	**0.758**	-0.062	-0.075	**0.318**	**-0.224**	**0.212**	**0.196**	**0.223**	0.153
	<.001	**<.001**	0.435	0.340	**<.001**	**0.004**	**0.007**	**0.012**	**0.004**	0.052

Note. TO = time online; ToSM = time on social media; PSES = perceived socioeconomic status; RSE = Rosenberg's Self-Esteem Scale; SAS-SV = smartphone addiction scale - short version; P.ECH = personal responsibility, high-severity scenario; O.ECH = someone else's responsibility, high-severity scenario; P.ECL = personal responsibility, low-severity scenario; O.ECL = someone else's responsibility, low-severity scenario; PoT = proportion of time online that is spent on social networking sites (presented separately as correlations were calculated using Spearman's *Rho*, whereas all other correlations presented are Pearson's *r*); values in bold are statistically significant at an alpha-level of .05.

As suggested in my correlation analyses, ToSM was the only substantial predictor variable when trying to explain participants' intentions to intervene, as measured by AUC. Other than for the O.ECH condition, $F_{O.ECH}$ (1, 161) = 2.96, p = .09, R^2 = .02, a model comprising of only the ToSM measure was significantly better than chance for predicting the rate of discounting across three of the study conditions, $F_{P.ECH}$ (1, 161) = 9.05, p = .003, R^2 = .05; $F_{P.ECL}$ (1, 161) = 3.92, p <.001, R^2 = .08; $F_{O.ECL}$ (1, 161) = 4.40, p = .04, R^2 = .03. The variance within the AUC values was explained by ToSM to a small degree, and the slopes from the regression models were modest ($\beta_{P.ECH}$ = .004; $\beta_{P.ECL}$ = .005; $\beta_{O.ECL}$ = .002). The proportional measure (PoT = Time on Social MediaTime Online) was also correlated with much of the discounting data, however, the distribution of PoT values was such that including it as a predictor variable was statistically inappropriate.

Study 5.2: Responsibility X Audience Size

Discounting Methods for Cyberbystander Research (RQ1 & H3)

The data presented in Table 12 show that participants in Study 5.2 discounted the benefit of intervening as the stated certainty in the scenario decreased (H5.3). Furthermore, that individual response sets were well-fit by Rachlin's (2006) equation (PL = 83%; OL = 90%; PS = 79%; OS = 86%) demonstrates the applicability of discounting methods for cyberbystander research (RQ5.1).

Participants rated that intervention would be most likely in the OL condition (Table 12). In other words, participants considered intervention most likely when it was someone else's responsibility and if a large audience was present, particularly when the certainty was moderate-to-high. The distribution of responses to the OL condition at high levels of certainty was also leptokurtic ($Kurtosis_{95\%}$ = 6.79; $Kurtosis_{90\%}$ = 5.11) compared to other conditions. Leptokurtic distributions are those that cluster close to the group mean. Thus, the distribution has less tailedness and usually has a smaller and less variable range of individual data points.

Put another way, participants responded most similarly when the certainty level was high and in the OL condition.

Table 12

Descriptive Statistics at each Certainty Value and Study Condition in Study 5.2

	95%				90%				50%				30%				10%				5%			
	PL	OL	PS	OS	PL	OL	PS	OS	PL	OL	PS	OS	PL	OL	PS	OS	PL	OL	PS	OS	PL	OL	PS	OS
M	76.69	87.85	77.45	81.15	76.84	83.23	74.83	78.97	54.43	56.71	51.56	53.50	43.66	42.90	44.64	40.58	33.81	32.31	34.01	30.30	29.47	26.23	30.92	26.57
SD	27.62	16.67	29.72	23.81	28.48	16.42	28.20	22.79	29.30	19.75	29.44	23.62	30.25	21.58	29.79	23.80	30.25	26.01	31.10	25.38	31.70	26.97	32.65	28.87
SEM	2.58	1.55	2.77	2.22	2.66	1.53	2.63	2.13	2.73	1.84	2.75	2.20	2.69	2.01	2.78	2.22	2.82	2.43	2.90	2.37	2.96	2.51	3.05	2.69
Mdn	91	95	90	87	88	86	85	83	55	57	54	53	40	40	40	37	21	26	24	22	15	19	17	15
Skew	-1.52	-2.23	-1.35	-1.93	-1.44	-1.70	-1.30	-1.97	-0.20	-0.25	-0.13	-0.32	0.31	0.43	0.16	0.48	0.86	1.00	0.70	1.03	0.93	1.27	0.81	1.04
Kurt	1.32	6.79	0.68	3.68	1.00	5.11	0.79	4.34	-0.91	0.64	0.79	-0.10	-0.79	-0.02	-0.99	-0.29	-0.39	0.25	-0.75	0.38	-0.40	0.81	-0.76	0.00

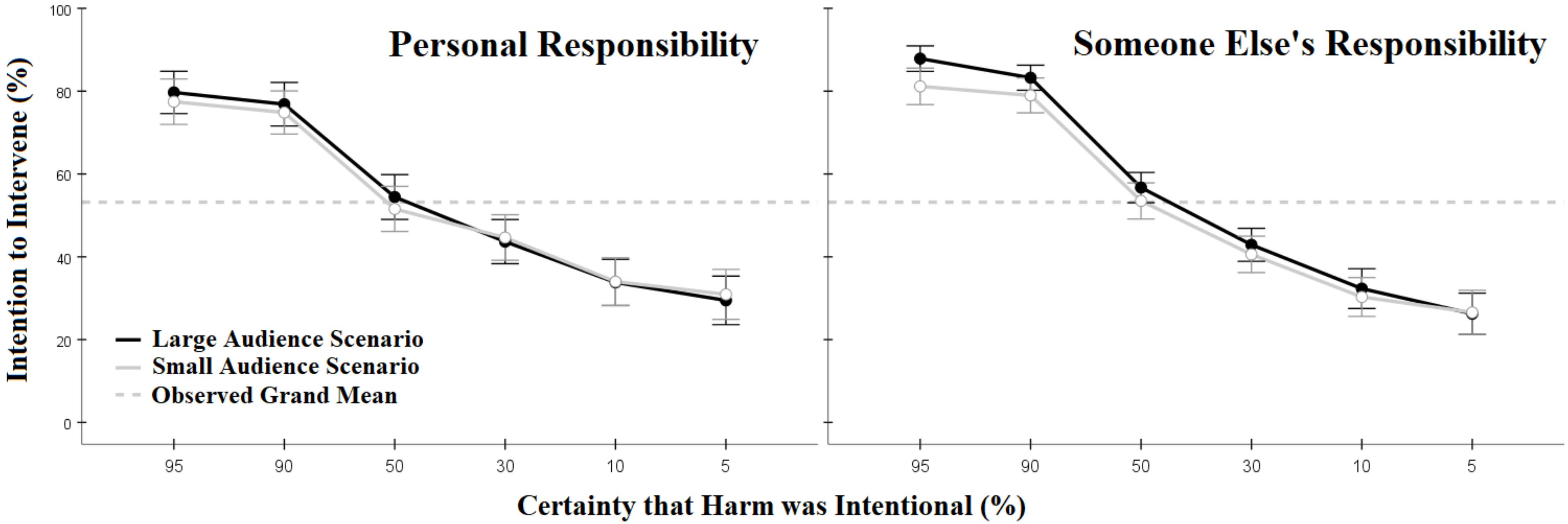
Note. 95%, 90%, 50%, 30%, 10%, and 5% are certainty values, *M* = mean indifference point, *SD* = standard deviation, *SEM* = standard error of the mean, *Mdn* = median indifference point, *Skew* = skewness, *Kurt* = kurtosis of the distribution of the indifference points (N = 163). PL = personal responsibility, large audience; OL = someone else's responsibility, large audience; PS = personal responsibility, small audience; O.ECL = someone else's responsibility, small audience.

Audience Size (H4) & Responsibility for Intervention (H5)

I tested H5.4 and H5.5, initially, using a repeated-measures ANOVA with Greenhouse-Geisser corrected values to account for raw response scores deviating from a normal distribution ($W = .62$, $p < .001$). I found no effect of study condition, $F(3, 342) = .90$, $p = .44$, $R^2 = .01$. I found a significant effect of the certainty variable (i.e., probability), $F(2, 185) = 431.69$, $p < .001$, $R^2 = .79$. A difference like this – between the raw responses as a function of the certainty level – adds to the evidence that participants discount the usefulness of intervening as a cyberbystander (i.e., RQ5.1). I also found a significant interaction effect between the conditions and the certainty values, $F(15, 1710) = 4.9$, $p <.001$, $R^2 = .04$. In Figure 8, I have plotted the aggregated intention-to-intervene raw responses from the PS and OL conditions as a function of the certainty that harm was intended. Most noteworthy in Figure 8 is that intention to intervene at high levels of certainty (i.e., 95% and 90%) was more likely when participants answered on behalf of someone else (i.e., Zee) and if a large audience was present.

Figure 8

Intention to Intervene as a Function of Certainty in Study 5.2



Note. Data are based on the raw response sets from participants and 95% confidence intervals are indicated by the error bars. Grey lines with open circles = the small audience size conditions; Black lines with closed circles = the large audience size conditions; the grey dotted line indicates the overall average intention to intervene.

As an alternate description of the intention-to-intervene data, I calculated the AUC for each participant's complete response set (Table 13). Across the four study conditions, the AUC values from the OL condition were highest (M = 0.57, Mdn = 0.57), least variable (SD = 0.17, SEM = 0.02, $Skewness$ = -0.10), and had the shortest-tailed distribution ($Kurtosis$ = 0.81). Regarding kurtosis in the distributions of intention to intervene across conditions, the data from the two someone-else's-responsibility conditions in Study 5.2 follow the same trend as the same conditions in Study 5.1. That is, data from scenarios that ask participants to take the view of another cyberbystander tend to be more tightly clustered around the average group-level response compared to when participants are answering according to their own likelihood of intervening. For example, the data from the OL condition in this study were leptokurtic ($Kurtosis_{OL}$ = 0.81) to a similar extent to the data of the O.ECH condition in Study 5.1 ($Kurtosis_{O.ECH}$ = 0.76). Further tests regarding the distribution of the AUC metric showed that data from all conditions met the assumption of normality ($K\text{-}S_{PL}$ = .05, p = .20; $K\text{-}S_{OL}$ = .06, p = .20; $K\text{-}S_{PS}$ = .05, p = .20; $K\text{-}S_{OS}$ = .07, p = .20).

Table 13

Area Under the Curve Descriptive Measures for each Condition in Study 5.2

	PL	OL	PS	OS
M	0.547	0.569	0.535	0.533
SD	0.259	0.174	0.264	0.200
Skew	-0.242	-0.096	-0.246	-0.230
Kurt	-0.560	0.810	-0.650	0.548
Mdn	0.545	0.570	0.552	0.532

Note. PL = personal responsibility, large audience; OL = someone else's responsibility, large audience; PS = personal responsibility, small audience; OS = someone else's responsibility, small audience.

I tested the effect of the responsibility-for-intervening manipulation (personal vs. someone else's) and the perceived audience size (small vs. large) using a 2-x-2 repeated-measures ANOVA. Neither of the main effects nor the interaction effect was statistically significant.

The combination of results generated using the raw response data and the AUC metric did not support H5.4. Specifically, participant responses in Study 5.2 did not significantly differ according to whether there was said to be a small or a large audience of others present. Furthermore, H5.5 was not supported. That is, how likely participants believed intervention to be did not significantly differ as a function of whose responsibility it was to intervene – a contrasting result to that of Study 5.1.

Fulfilment of the BIM Steps (RQ3 & RQ4)

Participants who indicated that they thought intervention was likely (i.e., those in the 50%-and-greater category) out-numbered those who rated intervention as comparatively less likely (i.e., those in the less-than-50% category). As shown in Table 14, most participants fulfilled Step 2 of the BIM (RQ5.3), in that approximately 55% considered that intervening was required across the four scenarios. The OL condition produced the highest proportion of participants who fulfilled Step 2 (62%). Therefore, recognising that a situation may require intervention (i.e., Step 2) was most likely to occur when participants knew there was a large audience of others present and that it was someone else's responsibility to intervene.

Table 14

Participants in Study 5.2 who Indicated that They, or Someone Else, Would Intervene

	PL	OL	PS	OS
Less than 50%	[a]51	44	54	50
	[b]44%	38%	47%	43%
50% and greater	64	71	61	65
	56%	62%	53%	57%

Note. [a]Frequency count of the participants who had an average indifference point in each category, e.g., less than 50% and in the PL condition. [b]Percentage of the sample that had an average indifference point in each category.PL = personal responsibility to intervene, large audience; OL = someone else's responsibility to intervene, large audience; PS = personal responsibility to intervene, small audience; OS = someone else's responsibility to intervene, small audience.

I also examined how often participant responses fulfilled Step 3 of the BIM (RQ5.4). Step 3 relates to whether a bystander takes personal responsibility for carrying out the intervention that - in the previous step - they considered to be necessary. The participants in Study 5.2 rated the intervention as more likely to occur when it was someone else's responsibility to intervene (Someone else's conditions: OL = 62%; OS = 57%), compared to when the scenario asked that they personally consider intervening (Personal conditions: PL = 56%; PS = 53%). As in Study 5.1, there was a discrepancy of 5% between those who personally took responsibility for intervening (54.5%) and those who thought someone else would intervene, although they would not (59.5%). Thus, 5% of participant responses fulfilled Step 2 of the BIM but not subsequently Step 3 (RQ5.4). That said, there was no statistically significant difference between the likelihood-of-intervention ratings as a function of whose responsibility it was to intervene in Study 5.2, $t(229) = -.50$, $p = .62$. A chi-square test of the across-conditions ratio of participants in the less-than-50%, versus in the 50%-or-more category, further supports that neither variable altered the likelihood-of-intervention ratings, $X^2(3) = 1.87$, $p = .60$, $\Phi = .06$.

Cyberbystander Intervention Covariates (RQ5)

Among the scale variables included in both Studies 5.1 and 5.2, there were several that were related across both studies. In Table 15, I have reported the correlation matrix in full. As was the case in Study 5.1 ($r = .298$, $p < .001$), TO was significantly correlated with ToSM (Study 5.2; $r = .494$, $p < .001$). ToSM was similarly positively correlated with SAS scores, both for participants in Study 5.1 ($r = .405$, $p < .001$) and those in Study 5.2 ($r = .315$, $p = .001$). Furthermore, the proportional measure of how much time online was explicitly spent on SNS (i.e., PoT) was positively correlated with SAS scores ($r_s = .245$, $p = .006$) and negatively correlated with participant age ($r_s = -.208$, $p = .026$), as in Study 5.1 ($r_s = .318$, $p < .001$; and $r_s = -.224$, $p = .004$, respectively). Regarding scale variables that correlated with discounting measures, only the scores from the PS condition were statistically related – and only to ToSM (r = .251, p = .007) and PoT (r_s = .222, p = .017). Given the absence of a scale variable that correlated significantly with most study conditions, I did not proceed to a regression analysis of the Study 5.2 dataset.

Table 15

Correlations Between Additional Measures and the Intention to Intervene AUC Values from Study 5.2 Conditions

	TO	ToSM	PoT	PSES	RSE	SAS	PL	OL	PS	OS
TO							0.042	-0.039	0.159	0.085
							0.656	0.679	0.090	0.366
ToSM	**0.494**						0.127	-0.055	**0.251**	0.083
	< .001						0.175	0.560	**0.007**	0.376
PoT	-0.072	**0.785**					0.149	0.020	**0.222**	0.084
	0.445	**< .001**					0.113	0.831	**0.017**	0.369
PSES	0.095	0.104	0.089				0.030	0.092	0.104	0.091
	0.312	0.268	0.345				0.752	0.327	0.268	0.332
RSE	0.044	-0.006	-0.032	0.093			0.071	0.075	0.098	0.023
	0.641	0.945	0.731	0.324			0.454	0.426	0.299	0.810
SAS	0.141	**0.315**	**0.254**	0.128	-0.148		-0.036	-0.050	0.025	0.010
	0.132	**0.001**	**0.006**	0.171	0.113		0.703	0.595	0.791	0.919
Age	-0.029	-0.142	**-0.208**	0.006	**0.184**	-0.128	0.049	0.095	0.023	0.025
	0.760	0.130	**0.026**	0.948	**0.049**	0.174	0.606	0.310	0.807	0.790

Note. TO = time online; ToSM = time on social media; PSES = perceived socioeconomic status; RSE = Rosenberg's Self-Esteem Scale; SAS-SV = smartphone addiction scale - short version; PL = personal responsibility, large audience; OL = someone else's responsibility, large audience; PS = personal responsibility, small audience; OS = someone else's responsibility, small audience; PoT = proportion of time online that is spent on social networking sites; values in bold are statistically significant at an alpha-level of .05.

Severity Ranking (H1 & RQ2)

As an extra part of Study 5.2, I conducted a Wilcoxon Signed-Ranks test to assess the validity of the low- and high-severity scenarios from earlier studies (H5.1). The rankings given by participants broadly matched how I had characterised these scenarios in Studies 4.1 and 5.1. The *PartyRudeness* scenario, the shortened version of the low-severity condition from prior work (i.e., ECL), was ranked moderately low in severity (M = 5.18, Mdn = 5). On the other hand, the scenario I employed as the comparatively more severe condition (i.e., ECH; *ExRunour*) was consistently ranked higher in severity (M = 6.51, Mdn = 7). I conducted a Wilcoxon Signed-Ranks test and found that the severity ranking applied to the low- (i.e., ECL) and the high-severity (i.e., ECH) scenarios differed in the expected direction, Z= 4.62, p < .001. Therefore, I found support for H5.1 in that participants' severity-ranking data validate that others perceive my initial low- and high-severity scenarios as such. Table 16 includes the shortened low- and high-severity scenarios and 18 scenarios with the average and median severity ranking that participants assigned them. Among all 20 short severity statements, it is worth noting that both the low-severity (i.e., ECL) and high-severity (i.e., ECH) scenarios were rated moderately severe at the group level.

Table 16

Severity-Rating Activity and Median Ranks

Comment in Full	Comment Identifier	*M* Rank	*Mdn* Rating
Y never receives any response from X when they try talking to them during online multiplayer games	IgnoredGaming	3.12	2
Y is not tagged in X's new picture online, even though they are in the picture and everyone else is tagged	TaggingError	3.20	2
Y receives a one-off mean Private Message from X	OnePM	4.12	3
Y finds out that they have been removed from a Group Chat without warning or explanation	ChatRemoval	4.17	3
Y's online Friend Request to X is ignored	FriendReq	4.17	2
Y knows that they are not included in the Group Chat that all their peers are part of	ExcludedGroup	4.52	3
X gets into a nasty argument with Y in a comments section, when a private message would have been less confrontational	ConfrontationalPM	4.90	5
On several different online posts of theirs, Y receives off-topic mean comments from X	OfftopicComm	5.08	5
Every time X posts a picture online, they mention Y in a mean hashtag	HashtaggedPhoto	5.15	5
X posted online to accuse Y of being rude to them at a recent party, even though Y was not present	**PartyRudeness**	**5.18**	**5**
X and Y have never gotten along well, then X unfairly posts online to tell all their mutual friends to start avoiding Y	InfluencingMutuals	5.25	5
Y is spreading untrue and hurtful information about X's recent break-up	RecentRumour	5.74	6
Every morning when they silently pass each other in the corridor, Y receives a mean Private Message from X	F2F_PM	5.79	6
X finds out the Y has been forwarding on their private messages, which they sent in confidence, to someone else	FwdPM	6.17	6
An ex-partner is spreading untrue and hurtful information about Y online	**ExRumour**	**6.51**	**7**
Y has sent explicit content to X several times, even though it is clear X finds it distressing	ExplicitContent	6.71	8
X posts an embarrassing video of Y online without their permission	EmbarrassingPhoto	6.79	8
X films Y without their knowledge, then posts the video online to make fun of them	VideoNonconsent	6.83	8
X gains access to Y's online account and pretends to be them in order to ruin Y's reputation	Impersonating	7.68	9
X and Y have recently broken up and Y lashes out by posting intimate content of X online to hurt them	ExSharing	7.70	9

Note. The two comments used (in their longer form) in my Pilot (4.1) Study and Study 5.1 are in bold font.

Further to validating the low- and high-severity scenarios from my earlier studies, I used the severity-ranking section of Study 5.2 to develop a scale version of the so-far dichotomous independent variable of severity level. Participants attributed a range of severity rankings across the set of 20 scenario statements ($M_{Minimum}$ = 3.12, $M_{Maximum}$ = 7.70). I ordered the statements in Table 16 according to their average severity ranking from participants. Statements considered least severe are at the top of the table (e.g., *IgnoredGaming*, *M* = 3.12, *Mdn* = 2), and those considered most severe at the bottom of Table 16 (e.g., *ExSharing*, *M* = 7.70, *Mdn* = 9). Ordered according to the aggregated average, severity rank aligns with the median severity rank for all but one statement: "Y's online Friend Request to X is ignored" (*M* = 4.17 and tied 4th-least-severe statement; *Mdn* = 2 and tied least-severe statement).

I include Table 17 as an alternate presentation of the severity-ranking activity dataset, as Table 16 includes the frequency that ranks applied to each scenario statement. Most statements received a single most-selected rank or two frequently selected consecutive ranks. For example, the *IgnoredGaming* statement was most often ranked first or second on the severity scale (1 = 23, 2 = 17), accounting for 68% of all rankings. For each statement, I have shaded the most frequently selected ranks in progressively darker grey, such that the most frequent rank is most grey and the least frequent ranks appear on a white background. Of note is the spread of the rankings of the *OfftopicComm* and *F2F_PM* scenario statements, which suggests these scenarios were the most variably interpreted. Participants more consistently ranked the remaining 18 scenario statements, and they are representative of the full severity scale (1-10), as was also shown by the aggregated data in Table 16.

Table 17

Frequency Count of the Severity Ratings Assigned to Each Scenario Statement

Rank	IgnoredGaming	TaggingError	OnePM	ChatRemoval	FriendReq	ExcludedGroup	ConfrontationalPM	OfftopicComm	HashtaggedPhoto	**PartyRudeness**	InfluencingMutuals	RecentRumour	F2F_PM	FwdPM	**ExRumour**	ExplicitContent	EmbarrassingPhoto	VideoNonconsent	Impersonating	ExSharing
1	23	21	7	6	23	10	2	2	1	3	1	0	2	1	4	5	4	6	5	4
2	17	7	12	16	10	9	4	9	4	10	3	7	1	3	6	3	2	3	2	2
3	3	7	14	9	2	13	10	5	9	8	9	6	8	5	6	6	4	2	1	3
4	2	3	7	9	1	4	7	4	12	32	9	5	9	7	6	3	3	1	4	2
5	4	2	5	3	2	3	12	8	8	21	12	7	7	7	12	3	5	5	2	2
6	1	0	1	3	2	3	6	9	8	22	11	6	7	11	25	4	6	2	3	0
7	1	0	2	3	1	3	2	7	9	19	8	14	9	13	21	7	4	4	2	1
8	0	4	5	2	2	5	5	4	8	8	5	6	5	9	27	9	7	6	8	5
9	3	3	3	5	6	6	3	2	0	6	2	6	4	7	15	8	13	16	15	7
10	5	2	3	3	9	4	1	2	1	1	1	1	5	3	8	15	10	9	23	24
% of N	45%	37%	45%	45%	44%	46%	40%	40%	46%	**99%**	47%	44%	44%	50%	**99%**	48%	44%	41%	50%	38%

Note. Low ratings (e.g., 1) indicate scenario statements that participants tended to rate as least severe; high ratings (e.g., 10) indicate comparatively high-severity scenarios; the most frequently selected rank or ranks (i.e., 1-10) for each statement are shaded grey. The *% of N* values is the percentage of the entire Study 5.2 sample who provided a severity rank. The statement identifiers in bold font are those that I adapted from studies presented earlier in this thesis.

The data from the severity-rating activity provided enough initial evidence to use the resultant ordinal scale of scenario severity in further studies. That said, I cautiously interpreted the severity scale results. The sample of participants who provided rankings was smaller than was ideal—with the exceptions of the shortened ECL and ECH scenarios presented to all participants, at most 65 participants contributed to the rating of any one scenario. The fewest participants rated the *TaggingError* and *ExSharing* scenarios (n = 49 and n = 50, respectively). The random sub-sample of scenarios that I gave participants to rate was important for managing the length of my study (i.e., 10 statements to rate rather than 20), though this meant that the sample used to initially validate each part (i.e., scenario statement) of my severity scale was small. Therefore, my subsequent use of this scale of scenario severity independent variable required further validation.

Gender Identity and Intervening

In some previous work, researchers have posited a possible relationship between a bystander's gender identity and their likelihood of intervening to help (e.g., Eagly & Crowley, 1986; Hoxmeier et al., 2020; Wolter & Napper, 2023). Regarding cyberbystanders in particular, Hayashi et al. (2023) investigated the relationship between social- and delay-discounting rates, participants' prior cyberbullying-related experiences, and whether either differed according to participant gender. They found mixed effects. For female participants, previous helping behaviours were related only to delay discounting, such that a higher rate of delay discounting correlated positively with having helped before. For male participants, previous helping behaviours were instead positively related to social discounting but not to delay discounting. Furthermore, Hayashi et al. (2023) found that their female participants reported having helped a target of cyberbullying significantly more than their male counterparts did.

To examine possible differences according to gender identity, I ran a 2-x-2 repeated-measures ANOVA and have presented these results in Table 18. Specifically, I tested the effects of scenario severity (low- and high-severity scenarios) and responsibility for intervening (personal and someone else's), as well as including gender identity as a between-subjects variable. In Study 5.1, the categories of female- and male-identifying participants exceeded 30 participants each and were the only groups appropriate for inferential analysis (n = 154). My results show that scenario severity and gender interacted, $F(1, 152) = 5.53$, $p = .02$, $R^2 = .01$. However, responsibility and gender did not, $F(1, 152) = .16$, $p = .69$, $R^2 = .00$. Results were consistent regardless of whether I used the response set averages (left column) or the AUC values (right column), as shown in Table 18.

Table 18

Statistical Relationship Between Main Effects and Participant's Gender Identity

		Response Set *M*			AUC		
Study	Variables	*F*	*p*	R^2	*F*	*p*	R^2
5.1	Severity	2.06	0.15	0.01	1.60	0.21	0.01
5.1	**Severity X Gender Identity**	**5.53**	**0.02**	**0.04**	**7.36**	**0.01**	**0.05**
5.1	**Responsibility**	**5.14**	**0.03**	**0.03**	**5.23**	**0.02**	**0.03**
5.2	Responsibility	0.58	0.45	0.01	0.78	0.38	0.01
5.1	Responsibility X Gender Indentity	0.16	0.69	0.00	0.08	0.78	0.00
5.2	Responsibility X Gender Indentity	0.47	0.49	0.00	0.48	0.49	0.00
5.1	**Severity X Responsibility**	**3.15**	**0.09**	**0.02**	**4.35**	**0.04**	**0.03**
5.1	Severity X Responsibility X Gender Identity	0.61	0.44	0.00	0.80	0.37	0.01
5.2	Audience Size	2.09	0.15	0.02	2.97	0.09	0.03
5.2	Audience Size X Gender	0.14	0.71	0.00	0.16	0.69	0.00
5.2	Audience Size X Responsibility	1.04	0.31	0.01	1.00	0.32	0.01
5.2	Audience Size X Responsibility X Gender Identity	0.00	0.98	0.00	0.00	0.95	0.00

Note. The degrees of freedom for Study 5.1 were 1 and 152 due to the reduced sample size of only the female- and male-identifying participants (n = 154). For Study 5.2 results, the degrees of freedom were 1, 109, and the reduced sample size was n = 111; values in bold are statistically significant at an alpha-level of .05.

Upon further examination, the significant interaction between gender identity and scenario severity was such that female-identifying participants rated intervention as more likely than did male-identifying participants in the high-severity conditions, as shown in Table 18. On the other hand, female- and male-identifying participants showed little difference in how likely intervention was when the scenario was low in severity. That I found some differences according to participant's gender identity – although inconsistent ones – does fit with the broader research in this area that is also mixed (Barlińska et al., 2013; Erreygers et al., 2016; Van Cleemput et al., 2014). As I did, Hayashi et al. (2023) found that female-identifying participants may be more likely to intervene to help others notwithstanding other factors (see also Bastiaensens et al., 2014; Macháčková et al., 2013). My findings build upon Hayashi et al.'s (2023) research because the discounting scenarios I used were explicitly about responding to aversive online behaviour rather than discounting rates correlating to separate questions on cyberbystanding.

Gender likely interacts with other relevant variables, which in turn influence cyberbystander behaviour (Hayashi et al., 2023). To this end, my results support the addition of an interaction between gender and responsiveness to scenario severity such that those who identify as female may, in the aggregate, be more likely to intervene in high-severity scenarios than they are in low-severity ones. In contrast, the intervention likelihood of those identifying as male was approximately equal regardless of the severity level.

As was the case when I used the full Study 5.2 dataset (N = 115), there also was not a main effect on AUC values according to the size of the audience within the male- and the female-identifying subset of participants, respectively, $F(1, 115) = 3.25$, $p = .07$, $R^2 = .03$, and $F(1, 109) = 2.97$, $p = .09$, $R^2 = .03$.

General Discussion of Chapter 5

Discounting Methods for Cyberbystander Research

The combined results from my pilot study (Study 4.1; Chapter 4) and Studies 5.1 and 5.2 presented in this chapter support my application of discounting methods to cyberbystander behaviour (RQ5.1). Specifically, my use of a probabilistic measure of how certain participants were that an aggressor intended to cause harm resulted in datasets that are typical of probability discounting. At high levels of certainty, participants across my studies were more likely to say that they would intervene to help or, when applicable, that they believed that someone else would intervene. As the interpretation of the situation as harmful became less clear (i.e., more ambiguity and less certainty), their likelihood of intervening decreased. Overall, then, I found support for H5.3. Additionally, in both Studies 5.1 and 5.2, the rate at which the likelihood to intervene decreased was well-described by Rachlin's (2006) hyperboloid discounting model. As other researchers have done, I interpret a well-fit discounting model as a strong indication that discounting methods apply to online social behaviour's less tangible social outcomes (e.g., Hayashi & Tahmasbi, 2020). Therefore, I found support for H5.3 and an answer for RQ5.1: Discounting methods are a useful tool for research into cyberbystander behaviour and how it functions given different environmental factors.

Traditionally, bystander behaviour has been studied in offline environments (e.g., Fischer et al., 2011). In the extant literature about bystander behaviour, researchers have documented that the likelihood of intervention to help a target of offline aversive behaviour can be diffused as a function of (1) the perceived severity of the scenario (Fischer et al., 2011; Sheeran et al., 2014); and, (2) the number of other people who are present (Abbate, 2022; Darley & Latané, 1968; Rabb et al., 2022). In terms of bystander behaviour occurring online (i.e., that of cyberbystanders), it has similarly been suggested that severity and the presence of others are important factors (Allison & Bussey, 2016; Macaulay et al., 2019). The results from my studies support that scenario severity impacts the likelihood that intervention will occur (H5.2, Study 5.1), yet the perceived audience size did not (H5.4, Study 5.2).

Scenario Severity

In Study 5.1, I reused the low- and high-severity scenarios from my earlier pilot study (i.e., ECL and ECH; Study 4.1) exactly as they were for the personal responsibility conditions (i.e., P.ECL and P.ECH, respectively). As was the case in the aggregate severity results from Study 4.1, the likelihood of participants personally intervening in the low- versus high-severity scenarios differed significantly. The difference – in both studies – was such that intervention was most likely when the scenario was high in severity, although marginally so (4% in Study 4.1 and 2% in Study 5.1, approximately). My results regarding the impact of scenario severity on intentions to intervene thus far align with others' results (e.g., Bastiaensens et al., 2014; Hayashi & Tahmasbi, 2020; Latané & Darley, 1970). In particular, Hayashi and Tahmasbi (2020) tested the impact of scenario severity and helping behaviour similarly to how I had, but with three levels of severity (mild, moderate, and severe). In that research, they suggest that helping behaviour is most likely – and that the rate of social discounting was the shallowest – when the scenario was severe (Hayashi & Tahmasbi, 2020). Worth noting, though, is that my high-severity scenario was moderately higher in severity than the low-severity scenario, as opposed to being of critical severity. It may be that the high-severity scenario in Study 5.1 is more analogous to Hayashi and Tahmasbi's (2020) moderate condition. Furthermore, my low-severity scenario could also be most analogous to Hayashi and Tahmasbi's (2020) moderate condition, which may explain the more marginal difference between the severity conditions in my study.

Given the margin of difference between my two severity scenarios and the possible discrepancy between my approach and that of others (e.g., Hayashi & Tahmasbi, 2020), I sought to validate that my moderately low- and high-severity scenarios were interpreted as such by others (H5.1). I developed the severity-ranking activity as an addition to the questionnaire for Study 5.2. Most participants in that study ranked shortened versions of the low- and high-severity scenarios, respectively, as moderately low in severity and moderately high in severity (Table 17). Furthermore, the average and median ranks that participants applied to the shortened low- and high-severity scenarios also validated their use as moderately low-severity and moderately high-severity scenarios (Table 16). In sum, I found support for H5.1.

I further explored scenario severity's impact on cyberbystander behaviour in the next chapter (Chapter 6) through analysis of the wider severity rankings (i.e., beyond the ECL and ECH scenario data). That said, preliminary descriptive analyses presented in this chapter indicated that the statements covered a range of severities and were, for the most part, consistently ranked among participants (RQ5.2).

Audience Size

Based on the work of Obermaier et al. (2016), in Study 5.2, I implied that there was either a small audience of other people present (small; approximately two others) or a moderately-large audience (large; approximately 50 others). In face-to-face aversive behaviour, an audience of other people may act as a protective factor for the aggressor in that bystanders are less likely to intervene to stop aversive behaviour (Darley & Latané, 1968; Fischer et al., 2011). In an online environment, though, the evidence for the effect of audience size is less clear (Allison & Bussey, 2016; Domínguez-Hernández et al., 2018). Some researchers have found that larger audiences do tend to inhibit the rate of cyberbystander intervention, although to a lesser extent than other contextual factors (Campos-Mercade, 2021; Hussain et al., 2019; Olenik-Shemesh et al., 2015). On the other hand, there is some indication that the presence of an audience has the opposite effect online–that an audience increases the perceived severity of the situation and, therefore, also the likelihood of a cyberbystander helping (Macaulay et al., 2022). However, a third possibility is that the ambiguity of the scenario – as viewed by the cyberbystander – eclipses any effect related to the size of the audience (Domínguez-Hernández et al., 2018). For example, that much SNS communication is asynchronous means that the audience of an online post continues to grow, making it difficult for cyberbystanders to perceive the audience size accurately. Based on results from Study 5.2, this third explanation – where ambiguity obscures the audience size – seems most likely, although further research is needed. Specifically, I found no difference in participants' likelihood of intervening as a function of the size of the audience; therefore, I did not find support for H5.2.

Responsibility for Intervention

I framed H5.5 bidirectionally due to the lack of clarity around whether prosocial intervention in aversive online behaviour functions primarily to garner positive or negative reinforcers for cyberbystanders. In both Studies 5.1 and 5.2, intention to intervene was higher in cases where I asked participants to consider what someone else would do, as opposed to what they themselves would do. However, the difference between the likelihood of participants versus someone else intervening – in the participant's estimation – was only statistically significant in Study 5.1. Of the individual study conditions, only the O.ECH (i.e., someone else's responsibility, high-severity scenario) condition was consistently responded to in a way that differed from the other Study 5.1 conditions. Overall, though, the results from across my two studies provide only weak support (Study 5.1) or no support (Study 5.2) for H5.5.

The BIM as Applied to Cyberbystander Behaviour

I considered the impact of whom participants believed to be responsible for intervening (themselves or someone else) to add to my exploration of how cyberbystander discounting data compares to the steps of the BIM (Latané & Darley, 1968). Specifically, I compared the proportion of responses where intervention was likely (i.e., 50% or greater) to those when intervention was less likely (i.e., less than 50%) across the personal and the someone-else conditions. In answer to RQ5.3, a majority of participants indicated that they considered intervention warranted in Studies 5.1 and 5.2 (60% and 57%, respectively). In regards to RQ5.4, fewer participants (although still a slight majority) indicated that they would take personal responsibility for intervening to help (55% and 52%, respectively). To summarise: across both Studies 5.1 and 5.2, there was a discrepancy such that approximately 5% of participants fulfilled Step 2 of the BIM but did not also fulfil Step 3 of the model. Although there is much research in which authors cite the BIM (e.g., Abbate et al., 2022; Butler et al., 2022; Ferreira et al., 2020), there are comparatively very few empirical measurements or tests of the earlier steps of the model (Mennicke et al., 2022). Thus, my exploratory analysis of the proportion of participants whose responses fulfilled Step 2 or both Steps 2 and 3 represents a novel contribution to the field of bystander research.

A further finding related to my third and fourth research questions (RQ5.3 and 5.4) came from the results of the OL condition in Study 5.2, where the proportions of those who fulfilled Step 2 of the BIM (62%), compared to those who did not (38%), was most marked. In other words, a larger proportion of participants interpreted that intervention was warranted (i.e., Step 2) when the scenario involved a large audience, compared to when I stated that only a few other people were present. As Macaulay et al. (2022) have proposed, that fulfilment of Step 2 was highest in my OL (Study 5.2) condition may be indicative that larger audiences increase the perceived severity of aversive behaviour in online environments. Future research should be designed specifically to test whether the factors of audience size, perceived severity, and the interaction of the two function differently online compared to offline.

Covariates

Of the additional measures I included across Studies 5.1 and 5.2, the most useful predictor/correlate of the likelihood to intervene measurement was the amount of time that participants spent on SNS (ToSM). In Study 5.1, participant ToSM was positively correlated with the likelihood of intervention in three of my four conditions. Additionally, ToSM was significantly correlated, in the expected direction, with most of the other additional variables: negatively with RSE, positively with SAS, and negatively with age. While other researchers have posited a relationship between self-esteem and the likelihood of cyberbystanders intervening to help targets (Macháčková et al., 2018; Schultze-Krumbholz et al., 2018), results from Study 5.1 suggest that it may instead be a relationship mediated by the amount of time spent on social media. Furthermore, participants' RSE scores were not related to their likelihood of intervening in either study or for any particular condition within Studies 5.1 and 5.2. Despite the theoretical connections between self-esteem and several of the BIM steps (e.g., McCarty et al., 2016; Paulhus & Martin, 1987; Witte, 1992), the idea that high rates of self-esteem are linked to an increased likelihood of prosocial cyberbystander behaviour (e.g., Lambe et al., 2017; Macháčková et al., 2018; Schultze-Krumbholz et al., 2018) was not supported by my results.

In both Studies 5.1 and 5.2, the additional ToSM and SAS measures shared a consistently moderate-strength correlation. That a self-reported estimate of time spent using social media (i.e., SNS) correlates positively with a scale intended as a measure of smartphone addiction adds to the evidence for such a scale's convergent validity (Alhassan et al., 2018; Kwon et al., 2013). That is, one component of smartphone addiction is likely to be how much time is spent accessing SNS using a smartphone (Scharkow, 2016; Schuster et al., 2022; Twenge et al., 2018); therefore, a correlation between ToSM and the SAS is expected. A correlation between two conceptually related constructs (i.e., time spent on a smartphone and smartphone addiction) supports that the SAS holds construct validity. However, the predictivity of self-reported time online and on SNS metrics should still be interpreted cautiously (Scharkow, 2016; Winskel et al., 2019). Moreover, research on the question of causality (or the lack thereof) between time on SNS and so-called smartphone addiction is mixed, and further research on the topic remains needed (Verduyn et al., 2017). Importantly, I did not find a difference between the SAS scores of participants recruited online and those who were recruited at a local University, suggesting that the SAS is appropriate for use with an Aotearoa New Zealand student sample.

Summary

Overall, the research in this chapter supports the use of probabilistic discounting methods to better understand how bystander behaviour functions in online environments. Using discounting methods, intervention was more likely to occur when the scenario was moderately high in severity rather than moderately low. I validated this so-far dichotomous severity variable and further developed it into a severity scale for use in a future study. Beyond the impact of scenario severity, the number of other bystanders present did not impact a participant's likelihood of intervening and was perhaps obscured by the ambiguity that is often a part of online interactions. Finally, the amount of time spent on SNS positively correlated with participants' likelihood of intervening, although how these variables relate requires further research.

Chapter 6: Severity Discounting & Cyberbystander Behaviour

Outline

In Chapter 6, I extend my thesis by using a scale version of the scenario severity variable. Studies 4.1 and 5.1 also involved a manipulation related to the severity of an online situation, though severity was a dichotomous rather than a scale variable in those earlier studies. I used extant literature (e.g., Bastiaensens et al., 2014; Hayashi & Tahmasbi, 2020) and a validation exercise (i.e., Study 5.2) to develop the severity scale I used as a novel independent variable in the present study (Study 6.1). The perceived severity of an online situation is an important contextual factor contributing to whether cyberbystanders intervene to help targets of aversive behaviour, and a more nuanced understanding of how severity and intervention interact is important for efforts to increase cyberbystanders' prosocial behaviour.

As well as scenario severity, I included the distinction between whether it was said to be the participant's responsibility to intervene (personal condition) or whether participants considered how another cyberbystander might behave (someone else's responsibility condition). As in Chapter 5, I used the responsibility variable to explore how often cyberbystanders fulfil Steps 2 and 3 of the Bystander Intervention Model (BIM; Latané & Darley, 1970). The BIM has been usefully applied to other types of bystander intervention and may also aid in the understanding of and ability to reinforce cyberbystander intervention.

In the latter parts of this thesis, I proposed (Chapter 3) and sought to validate (Chapters 4 and 5) the application of discounting methods to cyberbystander decision-making. Thus far, I have used a version of probability discounting in which the primary independent variable is the certainty (or lack thereof, i.e., ambiguity) that the aggressor intended to cause harm. Situational ambiguity (Latané & Rodin, 1969) and uncertainty regarding whether online actions were *intended* to cause harm (Olweus, 1997) – rather than inadvertently causing harm – are important factors in a bystander's calculus of whether to intervene. For online intervention, a related decision-making factor is how severe cyberbystanders perceive the situation to be (e.g., Bastiaensens et al., 2014; Latané & Darley, 1970; Macaulay et al., 2019). Perceived severity is the variable of focus in this chapter, both in terms of validating my previous results where I used a dichotomy of severity levels (Study 4.1, 5.1, and 5.2) and in extending these findings using a novel scale of perceived severity.

RQ6.1: Are discounting methods a useful tool for those researching bystander behaviour as it functions in the online environment?

Severity

The perceived severity of a situation has been an important factor for researchers of traditional bystander (e.g., Fischer et al., 2011; Janz & Becker, 1984; Phares & Wilson, 1972) and cyberbystander (e.g., Bastiaensens et al., 2014; Macaulay et al., 2019) behaviour. While there is some research on perceived scenario severity and cyberbystander behaviour, in-depth quantitative analyses of the interaction between severity and cyberbystander intervention have so far been limited (Bastiaensens et al., 2014; Zhao et al., 2023). My results from Studies 4.1 and 5.1 add to this literature and support the finding that cyberbystanders are more likely to intervene in scenarios perceived to be high in severity (Domínguez-Hernández et al., 2018; Latané & Darley, 1970). I designed the severity manipulation in my earlier studies (Studies 4.1 & 5.1) similarly to both Bastiaensens et al.'s (2014) and Hayashi and Tahmasbi's (2020) studies, in part to ascertain that results using indifference points (from discounting procedures) were comparable to those using other types of scale measurements. In both Studies 4.1 and 5.1, I explored whether different groups of participants responded in kind to the same low- and high-severity scenarios (relating to Hypotheses 4.3 & 5.2). As previous results were promising, I retained shortened versions of the original low- and high-severity scenarios (i.e., ECL and ECH) in Study 6.1 to explore further how different groups respond. Moreover, I contextualised scenarios ECL and ECH as two points on a scale of perceived severity to clarify the interaction between severity and intervening behaviour.

H6.1: The likelihood of intervening will increase the more severe the scenario is.

The results of Study 5.2 supported the hypothesis that participants appraised my dichotomous severity variable (the ECL and ECH scenarios) as such. I had conceptualised the categorical low- and high-severity scenarios as being either side of the mid-point on a severity continuum ranging from low to high because it is vital to study moderate and relatively ambiguous situations, as these are common and challenging (Macháčková et al., 2015; Marciano et al., 2022). I asked participants in Study 5.2 to rank the severity of the short versions of the ECL and ECH scenarios, in addition to a selection of similar statements (RQ5.2). The rankings from Study 5.2 participants validated my use of the ECL and ECH scenarios and formed the basis for my novel severity of online aversive behaviour (SOAB) 7-point scale. The SOAB fills a gap in the cyberbystander research literature; quantitative analysis of perceived severity has been limited (Bastiaensens et al., 2014), and of those quantitative studies, none I am aware of have included a scale (rather than categories) of perceived severity. I use the SOAB as a primary independent variable in Study 6.1 to assess how perceived severity relates to cyberbystanders' intentions to intervene. Participants of Study 6.1 also ranked the seven statements according to perceived severity as an exercise in validating the SOAB. Regarding the validity of the SOAB, I posited that:

H6.2: The 7-point Severity Scale ratings will align with the rankings from participants in Study 5.2 (i.e., will be validated across samples).

Perceived scenario severity also appears to be closely linked with Step 2 of the BIM (Koehler & Weber, 2018; Zhao et al., 2023), which is a necessary precursor to bystander intervention (i.e., Step 5 of the stepwise model; DeSmet et al., 2016). Step 2 in the BIM is that the scenario must be recognised as some form of emergency (Latané & Darley, 1970). Cyberbystander intervention appears more likely to occur when the severity of a scenario is high, perhaps because severe scenarios are more easily recognised as emergencies warranting intervention (i.e., fulfil Step 2; DeSmet et al., 2012; Obermaier et al., 2016). On the other hand, scenarios that are low in severity may be more difficult to appraise as needing intervention; therefore, Step 2 is less likely to be fulfilled in low-severity scenarios. Relatedly, some researchers have proposed that bystanders are more prone to blaming the targets of cyberbullying in cases where the perceived severity is lower, leading to greater reluctance to aid the targets (Koehler & Weber, 2018). In Studies 5.1 and 5.2, I calculated the proportion of participants who intended to intervene in 50% or more of cases where they had the opportunity to do so and considered these participants as having fulfilled Step 2. For both groups, most participants fulfilled Step 2 of the BIM regardless of study condition (Study 5.1 = 50-60%; Study 5.2 = 53-62%). I re-examined the proportion of participants who fulfilled Step 2 with the participants of Study 6.1.

RQ6.2: What proportion of participant responses fulfil Step 2 of the BIM?

Responsibility

Step 3 in the BIM relates to whether individuals take responsibility for intervening to disrupt the behaviour they recognised as an emergency (in fulfilling Step 2; Latané & Darley, 1970). Cyberbystanders deciding whether to take personal responsibility for intervening online are faced with competing positive and negative reinforcers (see Chapter 5 for a full review; e.g., Abbate et al., 2022; Crone et al., 2022; Schlund et al., 2021). When asked how likely someone else is to take responsibility for intervening, though, it is possible that potential consequences are less of a factor because the responder will not experience these (Allport, 1924; Jones & Nisbett, 1987; Weatherly et al., 1999). I began exploring whether cyberbystanders take responsibility for intervening in Studies 5.1 and 5.2. I posed a bidirectional hypothesis (H6.3) regarding whether personally intervening or someone else intervening was deemed more likely and found only weak support for the latter (and only in Study 5.1). Therefore, I used the Study 6.1 dataset to reexamine the same hypothesis and strengthen previous results.

H6.3: The intention to intervene will differ as a function of whether the responsibility for intervening is that of the participant or another person. If the intention to intervene is most likely in the someone else's responsibility condition, behaviour may be maintained by the avoidance of positive punishers. If the intention to intervene is most likely in the personal condition, behaviour may be maintained by trying to garner social reinforcement.

Specifically regarding whether Step 3 was fulfilled, I compared the intention-to-intervene rates for conditions in which participants answered on their own behalf (i.e., personal) and those in which I asked them to imagine what someone else would do. I used the same threshold as in earlier studies (i.e., intervening in 50% or more of opportunities to do so) to determine whether participants could be said to have fulfilled Step 3.

RQ6.3: What proportion of participants fulfil Step 3 of the BIM?

Severity X Responsibility

In addition to the main effects of perceived severity and whose responsibility it is to intervene (i.e., responsibility), these variables may also interact. For example, people might become less likely to personally intervene as severity increases, but their perception of whether someone else would intervene might not change across severity (e.g., Weatherly et al., 1999). Overall, the research regarding whether people make more impulsive (or riskier) decisions when deciding for themselves compared to when deciding for others is mixed (e.g., Neff & Macaskill, 2021). I asked participants how *they* would behave in the personal condition and how they thought *someone else* would behave in another condition. Neff and Macaskill (2021) found that people made more self-controlled (and less risky) decisions for other people and also when faced with a *should* question. In the context of my studies, whether a cyberbystander is likely or unlikely to intervene is more vital than whether either course of action is considered "self-controlled" or a risk. Nonetheless, the current discounting literature regarding risk and impulsivity is a useful starting point for examining the interaction of severity (i.e., level of risk) and a cyberbystander's own behaviour, compared to their perception of how others would respond (RQ6.4).

RQ6.4: How is the severity of an online scenario related to whether cyberbystanders take responsibility for intervening?

Additional Measures

The questionnaire for Study 6.1 included the same suite of additional measures used in prior studies (Studies 5.1 & 5.2; see Chapter 5 for a review). So far in my series of quantitative studies (i.e., Studies 4.1, 5.1, & 5.2), the additional variable that has emerged as the most useful predictor of cyberbystanders' likelihood of intervening is the amount of time they spend on SNS sites; that is, their ToSM score. My ToSM measure was positively correlated with intention to intervene in most Study 5.1 conditions (the exception being someone else's responsibility, high severity), as well as in a Study 5.2 condition (personal responsibility with a small audience). ToSM may be an important predictor of one's own intentions to actively engage online (e.g., Jensen et al., 2019; Valkenburg, 2021) but is less relevant when people make predictions about someone else's behaviour. Of the other additional measures I have included in this series, the SAS, designed to measure smartphone addiction, has shown good convergent validity (Kwon et al., 2013). My results thus far support its appropriateness for use with Aōtearoa New Zealand populations. Further, my results to date have not supported the idea that high self-esteem (as measured by the RSE; Rosenberg, 1965) is related to an increased likelihood of prosocial cyberbystander intervention (e.g., Lambe et al., 2017; Macháčková et al., 2018; Schultze-Krumbholz et al., 2018). My results to date also suggest no apparent gender difference in bystander behaviour, which is contrary to others (e.g., Hayashi & Glodowski, 2023; Hoxmeier et al., 2020). By including the same suite of additional measures across my studies, I can further explore these connections (or the lack thereof) between variables purported to be relevant for intervention behaviour.

> RQ6.5: How is the intention to intervene related to other variables (time online, time on social media, the proportion of online time that is on social media, smartphone addiction, perceived socio-economic status, and/or self-esteem)?

Current Study

With Study 6.1, I aimed to consolidate and strengthen several indicative results from my earlier work: that discounting methods can be applied to cyberbystander decision-making (Studies 4.1, 5.1, and 5.2); that cyberbystander decisions are influenced by perceived severity (Studies 4.1, 5.1, and 5.2); that a slight majority of participants fulfil Steps 2 and 3 of the BIM (Studies 5.1 and 5.2); and that ToSM explains part of the relationship between contextual factors (e.g., audience size) and intention to intervene (Studies 4.1, 5.1, and 5.2). Furthermore, I extended my previous work by including a new independent variable, a scale of short scenarios depicting aversive online behaviour of various levels of severity (the SOAB scale). Such a scale is imperative for furthering research on the relationship between perceived severity and the decision-making of cyberbystanders (Bastiaensens et al., 2014) –

cyberbystanders who are (perhaps more than aggressors or targets) an essential group for addressing aversive behaviour in the online environment (Allison & Bussey, 2016; Macháčková & Pfetsch, 2016; Ferreira et al., 2020).

Method

Recruitment

I recruited all participants from a population of undergraduate psychology students at an Aōtearoa New Zealand university. Many of the psychology papers available to these students included extra credit for research participation, which they could access independently of research and teaching staff on an online forum (SONA Systems, https://sona-systems.com). I posted a short description of my study on the SONA platform, available to students in October and November 2021. Students interested in taking part could click a link that took them to a consent form and, following that, the study questionnaire.

Sample

In all, 130 students began the Study 6.1 questionnaire. Of those, two only completed part of the survey. A further three took fewer than 5 minutes, less than the minimum and predetermined time required to complete the questionnaire satisfactorily. Another participant took longer than 45 minutes, which also fell outside the time range for satisfactory completion. An additional participant, who had completed the questionnaire in an adequate time, provided incorrect answers to 75% or more of the manipulation checks. Therefore, the final sample for Study 6.1 was 123 students.

Within the final sample (n = 123), there were slightly more 200-level-or-above students (52%) than 100-level students. On average, participants in this study were around 25 years old (M = 25.02, 95% CI [24.06, 26.05]). Age was not normally distributed ($K\text{-}S$ = .252, p < .001); however, this looked to be driven by a small proportion of participants who were 35 or older (n = 9), whereas most participants were in their early- to mid-twenties (20-25 =

88, 71.54%). I combined students of all levels into one sample because the groups were comparable across most measures. Of all the scale measurements, the 100- and 200+-level students only differed on RSE scores, $F(1, 122) = 9.70$, $p = .002$, $R^2 = .18$, and age, $F(1, 122) = 10.14$, $p = .002$, $R^2 = .18$.

Apparatus

All participants accessed the questionnaire from the SONA platform. The questionnaire was set up using Qualtrics and designed to be easily completed on any internet-capable device in the participants' own time.

Procedure

Discounting

Participants completed the discounting questions first, consisting of seven trials (severity levels) for each responsibility condition (14 trials in total; outlined in full in Appendix F). The seven scenarios of varying severity are in Table 19. Statements 3 and 5 are essential for later analysis because I had used versions of these previously (Studies 4.1, 5.1, and 5.2).

Table 19

Scenario Statements Used in Study 6.1

ID #	Scenario Statement
1	Y never receives any response from X when they try talking to them during online multiplayer games.
2	Y knows that they are not included in a Group Chat that X created and added all of Y's peers to.
3	X starts a nasty argument with Y in a comments section when a private message would have been less confrontational.
4	X posted online to accuse Y of being rude to them at a recent party, even though Y was not present.
5	An ex-partner (X) is spreading untrue and hurtful information about Y online.
6	X has sent explicit content to Y several times, even though it is clear Y finds it distressing.
7	X has filmed Y without their knowledge, then posted the video online to make fun of them.

Note. The scenario statements are listed from least- (Statement 1) to most severe (Statement 7) according to the ratings given by participants of Study 5.2.

The change in the level of responsibility was indicated by a short instruction on a page that preceded the respective conditions. The personal condition was preceded by: "Please answer the next seven questions according to how you personally would behave", and the someone-else condition by: "Please answer the next seven questions according to how you think someone else would behave, e.g., how an acquaintance of yours might behave. For the purpose of this section, imagine that this acquaintance is named 'Zee'." Each question within the personal condition started with "Imagine that you witness the following scenario online…", then included one of the scenario statements, then concluded with the prompt "How likely are you to protect Person Y by responding to call-out Person X?" The condition in which someone else was responsible for intervening (i.e., someone-else condition) began, "From Zee's perspective, imagine witnessing the following scenario online…" followed by one of the scenario statements, then "How likely do you think Zee is to protect Person Y by responding to call-out Person X?" I used gender-neutral pronouns and nondescript nouns (e.g., X, Y, and Zee) to refer to other people in the scenarios to minimise potential gender bias.

Across all the discounting trials, participants gave their response to each “How likely” prompt by moving the marker on a VAS that ranged from “Unlikely to reply, 0%” on the left through to “Will definitely respond, 100%” on the right-most end of the scale (e.g., Brumfitt & Sheeran, 1999; Kaplan et al., 2014). By default, I set the VAS marker at a neutral point (50%) to control for potential bias toward extremes of the scale.

I set up the Qualtrics questionnaire to randomise whether participants responded to the personal responsibility questions first or second. Within the personal responsibility and the someone else’s responsibility question sets, the severity statements were presented randomly to control for possible order and anchoring effects.

After each responsibility condition, participants answered two manipulation checks. The first manipulation check regarded whose perspective I had asked them to take (i.e., “In the previous seven questions, whose perspective were you being asked to take?”). The two response options were “Your own personal perspective” or “Another person’s (Zee’s) perspective. The second check was whether participants could correctly identify a distractor scenario statement among a list of statements included in the study condition they had just completed. I instructed participants, “From those below, please select the statement that did *not* appear within the previous seven questions”, where the distractor statement (and correct answer) was “Y’s online friend request to X is ignored”. The order of the options was randomised, and other than the distractor, the statements included were those in Table 19.

Additional Measures

In the latter half of the Study 6.1 questionnaire, I measured participants’ time spent in online activities (TO, ToSM, and PoT measures), self-esteem (RSE; Rosenberg, 1965), perceived social-economic status (PSES; Adler et al., 2000), the construct of smartphone addiction (SAS; Kwon et al., 2013), and basic demographic information (e.g., age; detailed in Appendix C). As an extra manipulation check and a further scale validation exercise (to build on Study 5.2), I asked participants to place the seven scenario statements from the discounting section in rank order according to their perception of each scenario's severity (activity shown in Appendix F).

Analysis

Having reviewed the literature on the different approaches to the analysis of discounting data (e.g., Bailey et al., 2021) and exploring these issues extensively in a pilot dataset (reviewed in Chapter 4), I did not exclude any data because they were nonsystematic or poorly fit by common models. In the first instance, I statistically analysed the raw indifference-point data and followed with analyses of the AUC values calculated from each response set.

Findings & Discussion

Discounting Methods for Cyberbystander Research

In Table 20 are descriptive statistics for the raw indifference-point likelihood of intervention data across both responsibility conditions (personal vs. someone else’s) and at all severity levels (1-7). Regardless of who was responsible for action, the likelihood of intervention was highest when the severity was greatest and decreased steadily as the scenario severity lessened, as hypothesised (H6.1) and as shown in Figure 9. The distributions of all indifference points fell within a range that indicates normality (based on skewness and kurtosis values), and, as further support, the standard error of the mean values was modest (*Range* = .73; from 1.88-2.61). The results from Study 6.1, like those of my earlier studies (4.1, 5.1, and 5.2), suggest that discounting methods are a useful tool for researchers of cyberbystander behaviour (RQ6.1).

Table 20

Descriptive Statistics for Each Severity Level in the Personal and Someone Else Responsibility Conditions

	Statement 7		Statement 6		Statement 5		Statement 4		Statement 3		Statement 2		Statement 1	
	Personal	S.Else	Personal	S.Else	Personal	S.Else	Personal	S.Else	Personal	S.Else	Personal	S.Else	Personal	S.Else
M	81.37	68.64	77.89	68.89	75.16	66.19	65.28	62.69	49.39	52.69	53.81	49.18	32.35	35.89
SEM	1.88	2.23	2.07	2.21	2.17	2.23	2.61	2.58	2.71	2.33	2.44	2.17	2.38	2.28
Mdn	84	73	81	72	80	70	70	69	52	60	55	50	27	34
Skew	-1.52	-0.67	-1.23	-0.67	-1.33	-0.69	-0.61	-0.68	-0.09	-0.35	-0.22	0.04	0.50	0.46
Kurt	2.73	-0.30	1.44	-0.24	1.55	-0.28	-0.62	-0.45	-0.97	-0.82	-0.83	-0.60	-0.56	-0.60

Note. The scenario statements are presented in reverse-severity order (i.e., right-to-left) to align with earlier displays where the scenarios most likely to engender intervention appeared first (on the left). *M* = mean, *SD* = standard deviation; *SEM* = standard error of the mean; *Skew* = skewness; *Kurt* = kurtosis; Personal = trials where participants answered on their own behalf; S.Else = trials where participants answered according to how they thought someone else would behave.

Severity X Responsibility

Mauchly's test for the assumption of sphericity was irrelevant for the two-level responsibility variable. For the scenario severity and the interaction analyses, the assumption of sphericity was relevant but was not met, $W_{Sev.}(20) = .543$, $p < .001$ and $W_{Int.}(20) = .563$, $p < .001$. For this reason, I report the Greenhouse-Geisser corrected values for all subsequent ANOVA results.

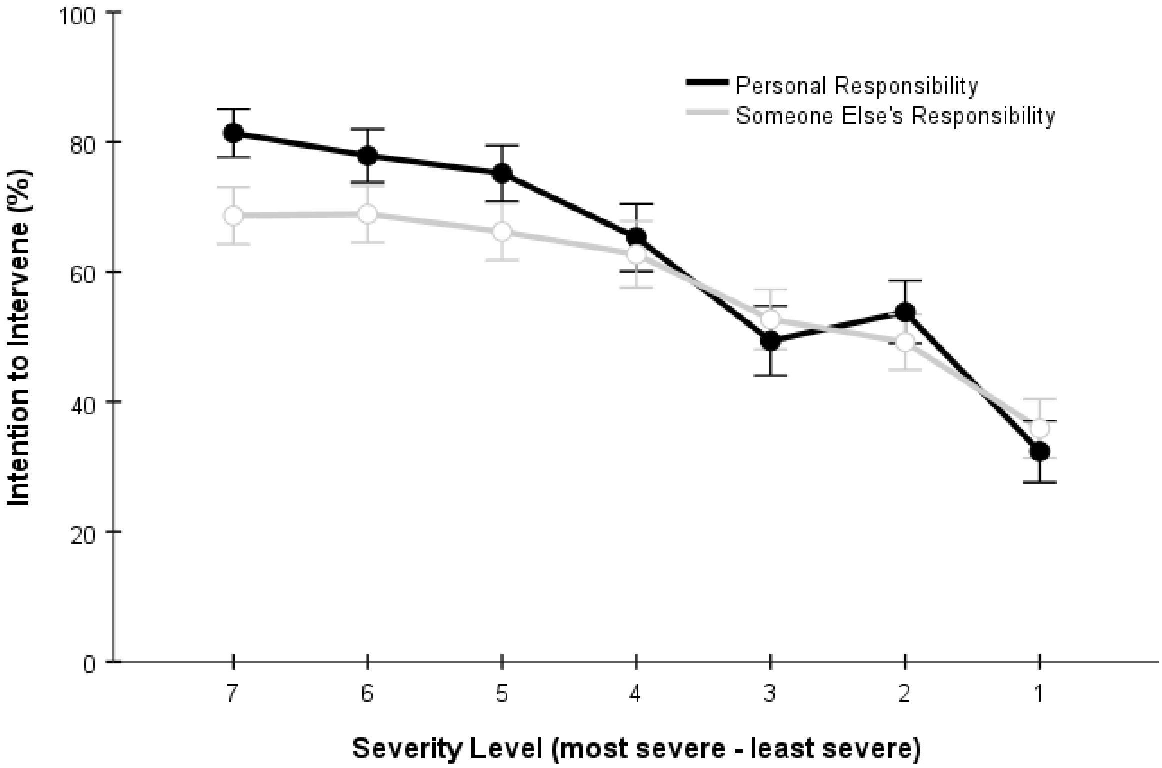
I found a small and significant main effect due to responsibility to intervene, $F(1, 122) = 6.28$, $p = .014$, $R^2 = .05$. On average, participants rated their own likelihood of intervening as higher than that of someone else ($M_{Personal} = 62.18\%$, 95% CI [59.02, 65.34]; $M_{S.Else} = 57.74\%$, 95% CI [54.59, 60.89]). In percentage terms, this suggests that people rate their intention to intervene as 4.4% (95% CI [0.6%, 8.1%]) more likely than that of others and supports H6.3.

There was an expected moderate-to-large difference between the various levels of scenario severity, $F(6, 732) = 101.99$, $p < .001$, $R^2 = .46$. That said, the pairwise comparisons revealed that while most levels of severity differed from the others, the three more severe statements were similar to each other. For example, there was little to no difference in the intention-to-intervene ratings between the most severe (Statement 7) and the next most severe statement (6), $M_{Diff} = 1.62$, $p = 1.000$, 95% CI [-3.37, 6.60]. In all, my results support H6.1, that the likelihood of intervening increases as the scenario severity increases.

In answer to RQ6.4, there was a small but significant interaction between responsibility and severity level, $F(6, 732) = 9.16$, $p < .001$, $R^2 = .07$. At higher levels of perceived severity, participants rated their own likelihood of intervening as higher than that of someone else. Conversely, at lower levels of perceived scenario severity, participants rated their likelihood of intervening as similar to that of others. Figure 9 shows the interaction between the most severe scenarios (left-most) and responsibility (personal = black lines, closed circles; someone else's = grey lines, open circles).

Figure 9

Intention to Intervene as a Function of Severity and Responsibility



The AUC values and descriptive statistics for each participant, both from the personal response set and the someone else's responsibility response set, are in Table 21. Using AUC, there was a significant effect on the likelihood to intervene between the responsibility conditions, $t(122) = 2.31$, $p = .023$, $R^2 = .08$. The difference was such that participants rated their own likelihood of intervening to help ($M = .622$) as higher than the likelihood someone else would intervene ($M = .578$), which further supports H6.3 and indicates the direction of the effect for future hypotheses.

Table 21

Area Under the Curve Descriptive Measures for Each Responsibility Condition

	Personal	S.Else
M	0.622	0.578
SD	0.177	0.175
Skew	-0.781	-0.417
Kurt	1.309	0.237
Mdn	0.640	0.584

Note. Personal = trials where participants answered on their own behalf; S.Else = trials where participants answered according to how they thought someone else would behave; *M* = average AUC; *SD* = standard deviation; *Skew* = skewness of the distribution; *Kurt* = kurtosis of the distribution; *Mdn* = the median AUC value.

As Statements 4 and 5 of the SOAB scale were, respectively, the low (ECL) and high (ECH) severity scenarios from Study 5.1, I have compared both sets of results in Figure 10. The Study 5.1 data presented are the average indifference points from each condition because participants gave several responses to each severity scenario as the primary independent variable in that study was the certainty of whether harm was intentional. The Study 6.1 data included in Figure 10 represents the participants' group-level indifference point for the two statements. I ran a 2-x-2-x-2 repeated measures ANOVA to determine whether intention-to-intervene differed at the study level (Study 5.1 vs. 6.1) and according to severity, $F(1, 121) = 44.15$, $p < .001$, $R^2 = .27$ and $F(1, 121) = 10.90$, $p = .001$, $R^2 = .08$. The same analysis showed no significant difference according to whose responsibility it was to intervene (personal vs. someone else), $F(1, 121) = .66$, $p = .42$, $R^2 = .01$. There were interaction effects between study and severity (low vs. high severity), $F(1, 121) = 4.14$, $p = .04$, $R^2 = .03$, as well as between study and responsibility, $F(1, 121) = 8.25$, $p = .01$, $R^2 = .06$.

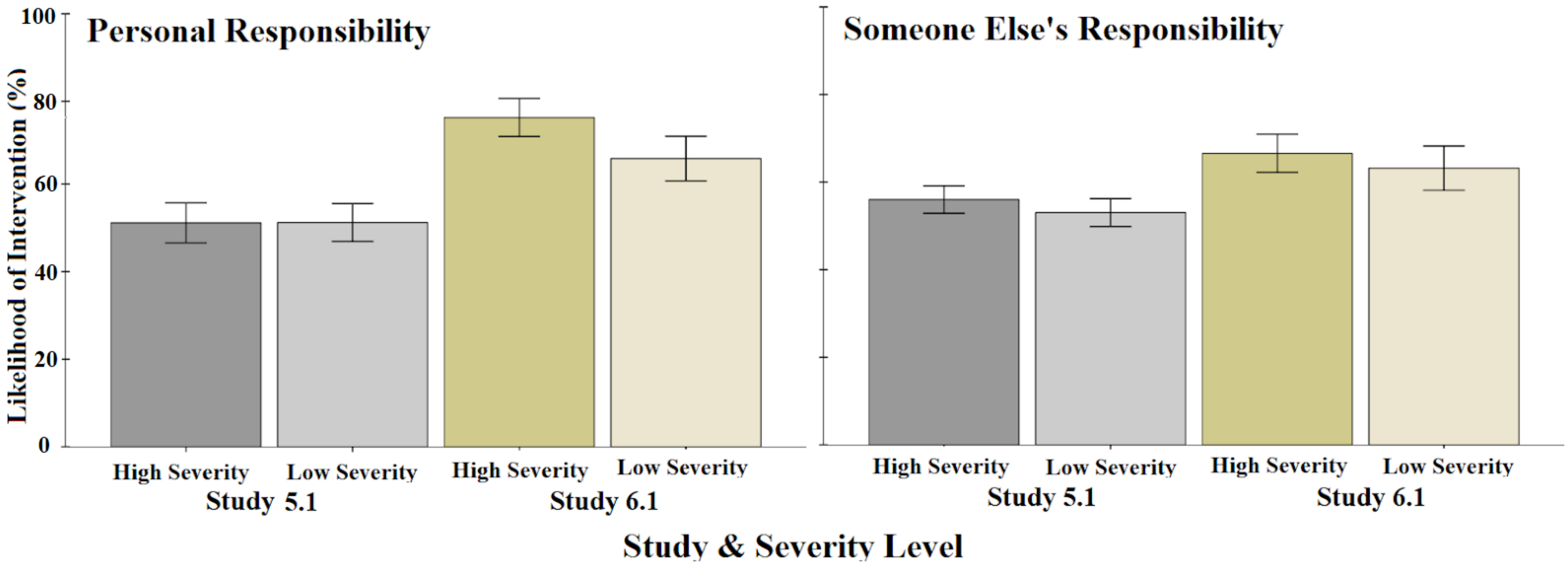
That there was a significant and medium-strength effect on intention-to-intervene values such that participants in Study 6.1 were consistently more likely to intervene than those in Study 5.1 is of note. The low- and high-severity scenarios were presented to the two groups in slightly different ways, which could have impacted how participants perceived them. In Study 5.1, participants read a short scenario that was several sentences in length. For example, the high-severity scenario was:

> Imagine that you are online when you come across a nasty rumour about a close friend of yours. A recent ex-partner of your friend has accused them of treating the ex badly and being wholly responsible for the breakdown of the relationship. The ex-partner has posted publicly and so the accusations are easily viewable to your friend, as well as all their friends and family.

Subsequently, in Study 6.1, I shortened these severity scenarios to one sentence. For example, the equivalent high-severity scenario was “An ex-partner (X) is spreading untrue and hurtful information about Y online”. Whether the changes in wording to the low- and high-severity scenario between studies are causally related to the higher intention-to-intervene scores in Study 6.1 requires follow-up research to clarify. Alternatively, participants of Study 6.1 may have been more sensitive to changes in the severity of the scenarios because they had additional statements, some of which were lower in severity and could have created an anchoring effect. That is, if participants had already responded to one of the lower-severity scenarios with a moderate intention to intervene prior to reading either of the PartyRudeness (Statement X) or ExRumour (Statement X) scenarios, they may have used their earlier responses as an anchor where there was less opportunity to do so in Study 5.1 given only two levels of severity.

Figure 10

Intention to Intervene in Sets of Low- vs. High-Severity Conditions



It is also possible that those who took part in Study 6.1 are different from those in Study 5.1 in some important way. The sample in Study 5.1 was made up of both University students and those recruited from an online research recruitment platform (Prolific.co). The Study 6.1 analyses are based on data from an entirely University sample. That said, there were no significant differences between the Prolific participants and the university students in the metrics collected as part of Study 5.1 (i.e., the discounting AUC values for each condition, duration, the extent of nonsystematic discounting, TO, ToSM, PoT, PSES, RSE, SAS, and age).

Fulfilment of the BIM Steps

Compared to my other work, Study 6.1 participants were likelier to say they would intervene. In both Studies 5.1 and 5.2, participants thought others (57 & 60%) would intervene more often than themselves (50 & 55%). In contrast, Study 6.1 participants responded that they personally were more likely to intervene (82%) than other people (70%), as shown in Table 22. Also noteworthy is that the proportions of Study 6.1 participants who would most often intervene (personal = 82%) or believed that others would most often intervene (someone else = 70%) was considerably higher than in my other studies (Studies 5.1 and 5.2). In sum, a greater proportion of Study 6.1 participants fulfilled Step 2 of the BIM (RQ6.2) compared to my prior studies (Studies 5.1 and 5.2). Similarly, more of the Study 6.1 sample fulfilled Step 3 of the BIM (RQ6.3).

Table 22

Number of Participants in Study 6.1 who Indicated that They, or Someone Else, Would Intervene

	Personal	Someone Else's
Less than 50%	22	37
	18%	30%
50% and greater	101	86
	82%	70%

Note. Frequency count of participants with an average indifference point of 50% or greater (lower panel) or an indifference point of less than 50% (upper panel). The frequency count for each category is presented as a percentage of the total sample.

In Studies 5.1 and 5.2, I observed a small discrepancy between the proportion of participants who seemed to fulfil Step 2 but who did not subsequently fulfil Step 3. In those studies (5.1 and 5.2), fewer participants indicated that they would intervene personally (i.e., would not take personal responsibility, Step 3) compared to the number of participants who believed that someone should intervene (i.e., recognise the intervention is needed, Step 2). There was no such discrepancy in the Study 6.1 dataset, evidenced by the higher likelihood of intervention in the personal-responsibility condition. As a starting point, this procedure of categorising intention to intervene values (e.g., fulfils vs. does not fulfil) to empirically test the early steps in the BIM is novel and fills an important gap within the broader bystander literature (Mennicke et al., 2022).

Cyberbystander Intervention Covariates

I included the same series of potential covariates in Study 6.1 as in Studies 5.1 and 5.2, which helped support the relevance of several variables related to cyberbystander decision-making (i.e., RQ6.5). Within the Study 6.1 dataset, several of these potential covariates were non-normal in distribution according to Kolmorogov-Smirnov tests (TO, $p = .05$; ToSM, $p < .001$; and SAS, $p = .03$). In addition, there is an argument for considering a single-item measure like the PSES as an ordinal scale, rather than a ratio one (Robitzsch, 2020). Therefore, I reviewed Spearman's coefficients in this instance to assess which variables were related and present the results in Table 23.

Participants reported spending 9-100 hours per week online ($M = 44.31$, 955 CI [41.13, 47.49], $Mdn = 42$), of which around 20 hours were spent on SNS ($M = 22.33$, 95% CI [19.79, 24.86], $Mdn = 19$, $Min = 1$, $Max = 77$). The PoT scores were normally distributed ($K\text{-}S = .06$, $p = .20$) and indicated that most participants spent approximately half of their online time specifically on SNS platforms ($M = 51\%$, 95% CI [47%, 56%]). TO and ToSM scores shared an equally strong negative relationship with RSE ($r_s = -.22$, $p = .014$) and a similar positive relationship with SAS (r_s = .41, p < .001 and r_s = .47, p < .001, respectively). Only ToSM scores were negatively correlated with age (r_s = -.31, p < .001), such that older participants reported less time on SNS. There were no significant differences between participant gender identity and other additional measures; importantly, this included no difference in time spent on social media related to participants' selected gender identity, $t(121) = .95$, $p = .34$, $d = .24$, $M_{Diff} = 3.37$ hours. ToSM has consistently been the best candidate for a predictor of cyberbystander intervention (see Chapters 4 and 5). In this case (Study 6.1), though, ToSM scores were significantly related to personal intentions to intervene (r_s = .21, p = .02), but not beliefs about others' intentions (r_s = .09, p = .33). Also of note was that age was negatively correlated with AUC values from the someone else condition, (r_s = -.19, p = .04). It may be that older cyberbystanders are less likely than their younger peers to intervene.

In terms of the next steps of my analysis plan (i.e., regression; Figure 6), ToSM was the best candidate for a predictive variable, at least for personal intentions to intervene. A regression model consisting of ToSM did explain 3.6% of the variance in participants' intention-to-intervene data, which was statistically significant, $F(1, 121) = 4.50$, $p = .04$, $R^2 = .036$.

Table 23

Correlations Between Additional Measures and Intention to Intervene AUC Values from Study 6.1

	TO	ToSM	PoT	RSE	PSES	SAS	Personal	S.Else
TO							0.159 0.079	0.109 0.232
ToSM	**0.516** **< .001**						**0.206** **0.022**	0.088 0.333
PoT	-0.136 0.135	**0.717** **< .001**					0.134 0.139	0.013 0.888
RSE	**-0.221** **0.014**	**-0.221** **0.014**	-0.084 0.355				0.128 0.158	0.121 0.182
PSES	0.027 0.767	-0.045 0.620	-0.108 0.234	**0.348** **< .001**			0.084 0.355	0.100 0.271
SAS	**0.414** **< .001**	**0.466** **< .001**	**0.225** **0.012**	**-0.346** **< .001**	-0.011 0.901		0.127 0.162	0.050 0.582
Age	-0.086 0.346	**-0.314** **< .001**	**-0.239** **0.008**	**0.205** **0.023**	-0.106 0.425	**-0.235** **0.009**	-0.105 0.249	**-0.185** **0.041**

Note. TO = time online; ToSM = time on social media; PoT = proportion of time online that is spent on social networking sites; RSE = Rosenberg's Self-Esteem Scale; PSES = perceived socioeconomic status; SAS = smartphone addiction scale - short version; Personal = study condition where participant responded according to what they personally would do; S.Else = study condition where participants responding according to what they thought someone else would do.

Severity Scale

In addition to including the newly developed SOAB scale as a primary independent variable (i.e., severity), I also asked participants to give their own severity ratings of each of the statements. One reason for doing so was as a validity check of the severity scale (H6.2). The second reason was to explore whether an individual's perception of the severity of a scenario can predict their likelihood of deciding to intervene (i.e., an extension to RQ6.4).

Severity Category & Scale (SOAB) Validity

As in previous studies, in the Study 6.1 data, I continued to find evidence that the original moderately low (ECL) and moderately high severity (ECH) scenarios are perceived as such by participants. In both the Study 5.2 and 6.1 questionnaires, I explicitly asked participants to give their perception of each statement's severity on a scale of 1-10 and 1-7, respectively (rather than only measuring the impact that statement severity had on subsequent cyberbystander-related decisions). For both cohorts, the shortened ECL statement (party rudeness) was rated as less severe than the ECH statement (ex rumour), $M_{Diff\ Study\ 5.2} = -1.90$, $t(120) = -13.43$, $p < .001$, and $M_{Diff\ Study\ 6.1} = -1.33$, $t(129) = -4.92$, $p < .001$.

For the SOAB scale overall, the perceived severity ratings given by participants who completed the Study 6.1 questionnaire were comparable to those from previous cohorts (i.e., Study 5.2 participants). To determine the comparability, I first collapsed the 10-point scale ratings down to a 7-point scale to directly contrast the ratings. Then I conducted a series of chi-square analyses of independence and report the results in Table 24. The chi-square results indicate that the two sets of severity ratings were statistically related. Therefore, I found support for H6.2, which also supports the validity of the SOAB scale items.

Table 24

Chi-Square Test Results and Descriptive Statistics for the SOAB Scale Validity

	X^2	*df*	*p*	Φ	*M*	*Mdn*	*Mode*
IgnoredGaming	25.40	5	< .001	.38	1.94	2	1
ExcludedGroup	41.34	6	< .001	.48	1.96	2	1
ConfrontationalPM	28.00	6	< .001	.40	2.66	3	3
PartyRudeness	21.20	6	.002	.29	4.56	5	4
ExRumour	36.29	6	< .001	.38	4.65	5	5
ExplicitContent	33.04	6	< .001	.42	5.57	6	6
VideoNonconsent	30.17	4	< .001	.46	6.66	7	7

Note. The left panel contains the chi-square results, including X^2 = Pearson's chi-square test coefficient, *df* = degrees of freedom for each test which differ in cases where not all ratings (i.e., 1-7) were selected, *p* = significance level, and Φ = the phi coefficient to indicate effect size. The right panel contains descriptive statistics of the severity ratings from the Study 6.1 dataset, including *M* = average severity rating, *Mdn* = median severity rating, and *Mode* = severity rating most often applied.

General Discussion of Chapter 6

Discounting Methods for Cyberbystander Research

In Chapter 6, I documented additional support for researchers' use of discounting methods to better understand cyberbystander decision-making, ultimately with a view to identifying initiatives that increase prosocial cyberbystander behaviour (RQ6.1).

Scenario Severity

Much of the bystander-related literature has shown that the severity of a situation is a key factor in determining whether someone will intervene to help (e.g., Bastiaensens et al., 2014; Fischer et al., 2006; Fischer et al., 2011). In my earlier studies, I found small but significant effects as a function of the situational severity such that the higher the severity, the more likely participants were to intervene (Study 4.1 = 2% more likely; Study 5.1 = 4% more likely). To extend upon those results, I developed a novel scale variable aimed at capturing different severity levels present in common online scenarios – the SOAB. Comparing intention-to-intervene scores with a scale measurement was useful because much previous research (including my own) compared results across a few discrete categories rather than against a continuum of severity (e.g., Bastiaensens et al., 2014; Hayashi & Tahmasbi, 2020). I used the severity ratings given by participants in Study 5.2, in conjunction with those from the manipulation check questions in this study (6.1), to confirm that my scale showed construct validity (H6.2). Using the SOAB scale, I found that intervention was most likely when the severity was high to moderately high and then steadily decreased as the severity lessened. Importantly, the decreases in intention to intervene between the different levels of severity tended to be statistically significant (H6.1).

Responsibility for Intervention

Across Studies 5.1, 5.2, and now 6.1, I included responsibility for intervening as a two-level independent variable. In the personal condition, participants responded according to whether they themselves would intervene. In the someone-else condition, participants responded according to whether they believed that someone else would intervene. I have so far posed a bi-directional hypothesis (H6.3) in regard to whether intervention is more likely in the personal condition or the someone-else condition due to contrasting theories (e.g., avoidance of becoming involved at all as self-protection, Schlund et al., 2021; a disconnect from the consequences when deciding for others, Weatherly et al., 1999), others' findings (e.g., Neff & Macaskill, 2021), and my own thus-far mixed findings. In contrast to Study 5.1, in which participants rated someone else as more likely to intervene, Study 6.1 participants rated their own likelihood of intervening as more likely than others (H6.3). In Study 6.1, I also found an interaction between responsibility and severity level such that participants rated themselves as more likely than others to intervene when the scenario was high in severity. At lower levels of severity, there was less of a difference (RQ6.4).

The BIM as Applied to Cyberbystander Behaviour

In this chapter, I have extended my work regarding the quantitative testing of the less-studied Steps 2 and 3 (as opposed to Steps 4 and 5) of the BIM (Mennicke et al., 2022). Prior to Study 6.1, I found that a slim majority of participants would fulfil Step 2, and slightly fewer (by approximately 4%) would fulfil Step 3. In Study 6.1, though, a clear majority (82%) of participants would fulfil both Steps 2 and 3 (RQ2 and 3). The difference in findings between studies may be due to untested factors present for one group of participants compared to the other. Another possibility is that the different results are partially explained by the change from discounting framed as a function of probability (i.e., the certainty that harm was intended; Studies 4.1, 5.1, and 5.2) to discounting framed as a function of severity in the present study (Study 6.1). Future replication-focused research using the SOAB scale should be undertaken to examine whether the intention-to-intervene results stay consistently high. If they are not consistent, further work hypothesising and testing the factors in Study 6.1 that lead to the comparatively high intention-to-intervene values is warranted.

Covariates

Among the variables that may be related to cyberbystanders' decisions about whether to intervene in online aversive behaviour or not, in Study 6.1, ToSM was linked only to personal intentions to intervene, while age was linked only to whether the participants believed that someone else would intervene (i.e., the someone else condition; RQ6.5). These two covariates were also related to each other in a way that suggests that ToSM decreased approximately linearly by the participant's age. Finally, although others have found gender differences in both discounting and SNS use (e.g., Hayashi & Glodowski, 2023), my results suggest no meaningful difference according to how a participant identifies.

Study 6.1 concludes the series of quantitative studies presented in this thesis, which I summarise in the next section (Chapter 7).

Chapter 7: Overall Discussion & Conclusion

Aversive Online Behaviour

Communication occurring in online environments often makes up a substantial proportion of our social behaviour (e.g., Andrews et al., 2015; Brinberg et al., 2021; Sun et al., 2023); it is, therefore, critical that online aversive behaviour is addressed and mitigated where possible. Young people are particularly vulnerable to the negative impacts of aversive online behaviour, including cyberbullying, because the use of social networking sites (SNS) is normative for this group (Boer et al., 2021; Kowalski et al., 2014). Moreover, if targeted by online aversive behaviour (compared to offline forms), young people may be more reluctant to seek help for fear that adults would not understand or that restrictions would be imposed on their own internet use (Bauman, 2015; Mishna et al., 2018; Pöyhönen et al., 2012). Adolescents I spoke to in a series of exploratory focus groups (Chapter 2) relayed several examples where the adults in their lives had not been able to support them regarding online social issues.

Notwithstanding the large range of prevalence estimates of online aversive behaviour that remain in the extant literature (e.g., Kowalski et al., 2019; Zhu et al., 2021), this behaviour does occur and has a detrimental impact on those targeted by it (Best et al., 2014; Dienlin et al., 2017; Kowalski et al., 2014). Thus, my focus for this thesis has been the processes and contextual factors that may influence social behaviour online. Principally, I was interested in bystander behaviour. Several focus group participants pointed to scenarios where support from bystanders was a key protective factor (Chapter 2), so I sought to better understand bystander behaviour, especially the conditions required for them to intervene in online aversive behaviour.

Bystander intervention has been found to help targets of aversive behaviour in many offline situations (e.g., Olweus, 1997; Pöyhönen et al., 2012; Salmilvalli, 2002). Further, though, my focus group participants and others suggested that the absence of support in open online environments might further compound the harm caused by the initial aversive action (e.g., Banyard, 2008; Bastiaensens et al., 2014; Macháčková & Pfetsch, 2016). Specifically, it seems as though one individual’s aversive behaviour toward a target is seen as representative of a larger group of people unless others (i.e., bystanders) show dissent. Furthermore, the actions of bystanders (who, by definition, are not directly involved in the negative behaviour) may be more amenable to change than those of the targets or perpetrators (Cassidy et al., 2013). Leung et al. (2018) have also documented that most people agree, at least in principle, that intervening to defend a target of aversive behaviour is the best course of action. The impact of bystander intervention cannot be understated (Van Heugten, 2011). However, research consistently indicates that many witnesses do not intervene in bullying incidents or other aversive behaviour (Craig & Pepler, 1998; Huang & Chou, 2010; Schultze-Krumbholz et al., 2018). Therefore, I investigated the applicability of behaviour analytic approaches to the social problem of cyberbystander behaviour (i.e., bystanders who are in online spaces).

Behaviourist perspectives highlight the environmental influence on human behaviour, with social outcomes such as reinforcement or punishment playing a crucial role (Baum, 2017; Skinner, 1953). SNS represent online social environments where harmful behaviours can manifest (Lysentøen et al., 2021). Bullying and related forms of aversive behaviour are often maintained by social reinforcement. The presence and behaviour of bystanders in the online environment play a significant role in shaping aversive behaviour because it tends to be their explicit or implicit support of the aggressor that reinforces and therefore perpetuates such behaviour (e.g., joining in with the aggressor or failure to interrupt the aversive behaviour, respectively; Ross & Horner, 2009; Salmivalli, 2002). On the other hand, bystander intervention to help the target removes the social benefit for the aggressor (i.e., negative punisher), which should, in turn, reduce the likelihood of future aversive behaviour while also showing critical support for the target. In many cases, there is no reason to expect that online aversive behaviour is controlled by contingencies that differ from those that control offline aversive behaviour. For example, the greater levels of anonymity afforded by the online environment can be cited as a unique difference between what influences online compared to offline behaviour (e.g., Sticca & Perren, 2013; Wright, 2013). However, online aversive behaviour also occurs in spaces where people are identifiable (e.g., Facebook; Burke et al., 2011; Uram & Skalski, 2022), and simply because a space is anonymous, it does not necessarily follow that people behave poorly (e.g., Small & Loewenstein, 2003). There are certainly exceptions, such as in cases where the physical distance between an aggressor and target becomes a critical factor. I worked from the premise that explanations of offline aversive and bystander behaviour are a useful starting point in understanding the respective online behaviour.

The Bystander Intervention Model

The Bystander Intervention Model (BIM) was originally formulated by Latané and Darley (1970), the former of whom became interested in the idea of bystander behaviour following the coverage of the murder of Kitty Genovese in the media. Although it is now best known as the genesis point for the bystander effect literature, there is a strong argument for Kitty Genovese's story being far more emblematic of media and the authorities misrepresenting events (Manning et al., 2007), of bystander empathy (rather than apathy; Kassin, 2017), and of making space for joy and passion even in otherwise hostile environments (e.g., 1960's New York for the queer community; Hobbes & Marshall, 2019). That argument deserves a thesis of its own and is beyond the scope of this one; my argument instead begins at the point where Darley and Latané's (1968), as well as their successors' (e.g., Fischer et al., 2011; Latané & Nida, 1981; Latané & Rodin, 1969), work supported that there was a "bystander effect" when it came to responding to emergencies (although, not in

Kitty Genovese's case). While not developed within the framework of behaviourism, the BIM (Latané & Darley, 1970) is a step-wise model where the steps can be likened to a chain of behaviours that culminate in bystander intervention (Baum, 2017). The BIM has five steps (see Figure 1). Because the BIM is cited as being stepwise (e.g., Fischer et al., 2011; Latané & Darley, 1970), research exploring the early steps of the model should be just as important as testing Step 5, which is the act of intervening. That said, there has so far been limited empirical work testing Steps 2 and 3 of the BIM (Mennicke et al., 2022). More generally, the BIM has been successfully applied to behaviour change initiatives relating to intervening in various forms of offline behaviour (e.g., Lassiter et al., 2021; Worsteling & Keating, 2022). Recently aspects of the BIM have also been applied to online behaviour (Abbate et al., 2022; Bastiaensens et al., 2016; Kazerooni et al., 2018). One option for assessing how people (such as cyberbystanders) behave in instances of aversive behaviour is to pose hypothetical scenarios and ask people to imagine and report how they would respond (Bickel & Marsch, 2001). As a self-report measure, it is possible that this method will overestimate the prevalence of prosocial behaviour (Lanz et al., 2021), such as helping targets. To mitigate this possibility, I used discounting methods where participants responded multiple times to scenarios, each with small differences (e.g., differences in the severity level in the case of Study 6.1).

In addition to investigating the impact that various contextual factors have on cyberbystander decision-making, I designed my discounting studies to measure the fulfilment of the early steps of the BIM as a means to begin assessing the applicability of the model to online behaviour. In Studies 5.1 and 5.2, approximately half of the participants fulfilled Step 2 (i.e., recognised the scenario as harmful and warranting intervention), and about 5% fewer fulfilled Step 3 (i.e., would take responsibility for intervening). In Study 6.1, a higher proportion of participants fulfilled both Steps 2 and 3 (82%). The observed variation in fulfilment rates in Study 6.1 may be partly attributed to the shift from discounting framed in terms of probability certainty, as seen in Studies 4.1, 5.1, and 5.2, to discounting based on the concept of severity in the present investigation (Study 6.1). That a change in the framing (i.e., certainty vs. severity) of the online aversive behaviour could be reflected in distinctly different intention-to-intervene values underscores the nuanced role that contextual factors play in shaping cyberbystander decisions.

Applying Discounting Methods

The application of discounting methods has proven successful in various decision-making situations, particularly those involving impulsive or risk-laden choices (e.g., Andrews, 2022; Dixon et al., 2006; Hayashi et al., 2016). These scenarios most often require individuals to weigh immediate benefits against potential larger or uncertain consequences (Poltavski & Weatherly, 2013; Tian et al., 2018). Cyberbystanders deciding whether to intervene in instances of online aversive behaviour fit within a discounting framework. For example, there may be some immediate risk that the aggressor will turn their attention to the cyberbystander, but there are larger and possibly more enduring impacts for the target of the aversive behaviour if no one intervenes to help them (Dillon & Bushman, 2015; Sticca & Perren, 2013). Moreover, there are also large but perhaps less tangible impacts on what kinds of online social behaviour are considered normative (Bastiaensens et al., 2016; Macháčková & Pfetsch, 2016) if poor behaviour is not punished – in the behavioural sense. Through a series of studies (4.1, 5.1, 5.2, & 6.2), I investigated whether discounting methods could be applied to cyberbystander decision-making.

Probability Discounting

In my first three studies (4.1, 5.1, and 5.2), I specifically applied probability discounting. Probability discounting is a measure of whether and to what extent an individual devalues an outcome under different levels of risky (i.e., uncertain) circumstances (Green et al., 1999; Reynolds et al., 2004). According to bystander effect research, the level of situational ambiguity (Phares & Wilson, 1972; Solomon et al., 1982) in a scene significantly impacts a bystander's likelihood of intervening to offer assistance (Darley & Latané, 1968). In Studies 4.1, 5.1, and 5.2, I framed the level of situational ambiguity present in a hypothetical scenario depicting online aversive behaviour in terms of probability. I used the probability that the aggressor intended to cause harm to their target to investigate how a cyberbystander's perception of intentionality (i.e., a criterion of bullying; Olweus, 1997) impacted their likelihood of intervening. Put another way, I used certainty (of aggressor intentionality) as the primary independent variable and measured how likely cyberbystanders were to intervene as the dependent variable. I asked participants the broad question, "Would you intervene?" The gradients of the resulting discounting curves represent the extent to which participants' initial preference of whether to intervene (at high certainty levels, i.e., 95%) was devalued as the situational ambiguity increased (as the certainty lessened). For example, steep probability discounting in this context suggests that uncertainty about the nature of an online social interaction disproportionately reduces cyberbystanders' intention to intervene.

Most existing discounting research has primarily focused on commodity-based outcomes, such as tangible monetary rewards (e.g., Small & Lowenstein, 2003). However, given that aversive behaviours, including bullying, are often sustained through social reinforcement or punishment (Salmivalli, 2002), it was imperative that I consider social outcomes in understanding cyberbystander behaviour. In my pilot study (Study 4.1), I included a pair of scenarios with monetary consequences and compared these against a pair of scenarios with the kind of social consequences commonly associated with online behaviour. Social-consequence conditions showed a similar pattern to monetary ones. Therefore, I found support for applying discounting techniques to situations involving social consequences related to cyberbystander behaviour.

The initial results from my pilot study (Study 4.1) supported the premise that social outcomes related to cyberbystander behaviour are discounted as a function of probability in a way that is similar to existing applications of discounting (Smith & Hantula, 2008). Results from Studies 5.1 and 5.2 corroborated the idea that cyberbystander outcomes were discounted similarly. At the aggregate level across the studies and conditions, participants exhibited high intention-to-intervene values when their certainty was high (i.e., 95%). As the certainty around whether harm was being perpetrated intentionally lessened, so too did participants' likelihood of intervening (i.e., intention-to-intervene values also lessened). In sum, my studies supported the assumption that probability discounting can be usefully applied to cyberbystander decision-making.

When using discounting methods, there are some established guidelines that many researchers follow (e.g., Johnson & Bickel, 2008; Smith & Hantula, 2008). It is important to investigate whether these standard guidelines continue to apply when discounting is applied to a novel topic (Johnson & Bickel, 2008). I found that adhering strictly to the guidelines could lead to an underestimation of participants' intentions to intervene in cyberbystander situations. Thus, I chose not to exclude any response sets based on failure to meet Johnson and Bickel's (2008) guidelines. Instead, I considered these nonsystematic response sets as valuable indicators of cyberbystander behaviour rather than mere methodological anomalies (Ballard et al., 2023). Most crucially, participants whose responses suggest they would always or almost always intervene (i.e., *always-interveners*) produced what would be considered nonsystematic discounting data. Although the focus of my research was to examine the conditions that most impacted the likelihood of intervention (e.g., at what severity level does likelihood lessen?), these always-interveners are also important to examine. Across my studies, always-interveners had marginally lower PSES and RSE scores (see Table 25) but were otherwise very similar to the general participants on the additional measures. To increase the overall incidence of cyberbystander intervention, it would be useful to determine what distinguishes always-interveners from others, as this may highlight an avenue for behaviour change initiatives. My findings suggest that a considered rejection of the standard discounting guidelines is sometimes warranted.

Table 25

Additional Measures for Always-Interveners versus General Participants

	Always-Interveners		General Participants	
TO	51.61	[47.94, 55.28]	47.88	[46.95, 48.82]
ToSM	25.99	[23.22, 28.75]	23.77	[23.03, 24.51]
PoT	56.35%	[51.68, 61.01%]	53.21%	[51.67, 54.75%]
PSES	5.43	[5.19, 5.68]	6.10	[6.03, 6.17]
RSE	20.39	[19.35, 21.44]	24.14	[23.84, 24.44]
SAS	29.85	[28.71, 30.99]	30.25	[29.79, 30.72]
Age	26.27	[24.95, 27.60]	25.39	[25.05, 25.73]

Note. The overall average (and 95% confidence intervals shown in square brackets) additional measure values for participants categorised as always-interveners (left panel) compared to all other participants (right panel), where participants from Studies 4.1, 5.1, 5.2, and 6.1 are combined. Always-Interveners are those who consistently responded with 90% or greater intention to intervene regardless of either the certainty or the severity. TO = time spent online; ToSM = time spent on SNS; PoT = proportion of online time that is spent on SNS; PSES = perceived socioeconomic status; RSE = Rosenberg's self-esteem scale; SAS = smartphone addiction scale.

Existing research has indicated that incident severity is linked to the likelihood of bystander intervention online (Bastiaensens et al., 2014). Specifically, it has been found that as the severity of an incident increases, bystanders become more likely to intervene (e.g., Fischer et al., 2011; Phares & Wilson, 1972). Additionally, Sheeran et al. (2014) reported a medium-to-large effect of risk appraisal on bystanders' resultant behaviour, emphasising the importance of perceived incident severity. However, independent testing on the effectiveness of perceived incident severity intervention strategies has been limited (Koehler & Weber, 2018). To that end, I included a dichotomous severity-level independent variable in both

Studies 4.1 and 5.1, which showed that cyberbystanders are more inclined to intervene in scenarios perceived as high in severity, in line with other research findings (Domínguez-Hernández et al., 2018; Latané & Darley, 1970). In Study 5.1, the difference in responses between high- and low-severity scenarios was such that intervention was 2% more likely when the severity was high. Studies 4.1 and 5.1 contribute to the existing literature on bystander intervention, providing further evidence of the importance of perceived incident severity in shaping cyberbystander behaviour, and I extended this research in Study 6.1 (Chapter 6).

It is likely that there are complex series of contingencies that influence bystander behaviour (Kowalski et al., 2014; Latané & Nida, 1981). Active bystanders in online environments may face the risk of retaliation from the aggressor in response to their interference, as documented by Abbate et al. (2022). Intervening in defence of others exposes bystanders to the possibility of becoming new targets for aversive behaviour, serving as a positive punisher of their intervention efforts. Conversely, bystanders may choose to avoid intervening altogether, removing themselves from the chance of being harmed, which acts as a negative reinforcer (Dymond et al., 2014; Schlund et al., 2021). Another motivating factor for cyberbystander intervention could be the pursuit of positive social reinforcement and a sense of being morally superior. Codol (1975) proposed that intervening to help others allows bystanders to reinforce their self-perception as good or superior individuals. If individuals perceive themselves as better than the average person, intervening may be congruent with their self-image and represent the option with the highest reinforcing value (Alicke et al., 1995; Brown, 2012). Consequently, participants might rate themselves as more likely to intervene compared to others.

To begin examining the contingencies that may help explain bystander behaviour, I manipulated whether participants answered about their own likelihood of intervening (personal condition) or about the likelihood of someone else intervening (someone else's responsibility condition). If intervention is more likely when participants answer on their own behalf, explanations such as that cyberbystanders are pursuing positive reinforcement may be most accurate. Conversely, if intervention is less likely when participants answer on their own behalf, explanations like the avoidance of harm and retaliation should be favoured and further tested.

In Study 5.1, participants rated their own likelihood of intervening at 52.9%, which was lower than their perception of how likely someone else would be to intervene at 56.4%. However, in Study 5.2, I found no significant difference between participants' own likelihood of intervening and their perception of someone else's likelihood of intervening. In Study 6.1, participants rated their own likelihood of intervening as more likely than others, with a percentage difference of 4.4%. That my findings were mixed suggests that there is a complex interplay between perceived risks, moral considerations, and beliefs about the prevailing social norms (i.e., what someone else would do). A future study in which participants are asked both what they believe “should” be done versus what they “would” do may help clarify these *responsibility* results. By framing the question to participants as “what would you do?”, I would expect participant behaviour to be more influenced by the perceived personal risks of intervening, as compared to responses to the prompt “what should you do?” (Neff & Macaskill, 2021). Using “what should you do?” in place of “what would someone else do?” also may more clearly represent participant views of the social norms regarding cyberbystander intervention.

Within my probabilistic discounting studies (i.e., 4.1, 5.1, and 5.2), I also investigated the impact that the audience size had on cyberbystander intervention behaviour. In research about bystander behaviour, it has been documented that a large audience inhibits any individual audience member (i.e., bystander) from intervening (e.g., Kazerooni et al., 2018;

Latané & Nida, 1981). One idea regarding why intervention is less likely in cases where there is a larger audience is that people are less likely to recognise that intervention is required. Put another way, a large audience of other people who are not intervening implicitly gives the impression to each individual audience member that intervening is not required (e.g., Butler et al., 2022; Phares & Wilson, 1972). Furthermore, even if individual audience members recognise that intervention is required (i.e., Step 2 of the BIM), they may think that someone else will intervene, and therefore, they do not assume responsibility for intervention (i.e., Step 3 of the BIM). In the extant bystander effect literature, researchers refer to this as a "diffusion of responsibility" (Darley & Latané, 1968).

I did not find evidence of a diffusion of responsibility brought about by a larger audience size in Study 5.2. In my study, I explicitly informed participants that the audience of other bystanders was either two or 50 people. In a naturalistic online setting, though, it is often unclear precisely how many other people witness an instance of aversive online behaviour (e.g., Buckley et al., 2022; Domahidi, 2018; Zhu et al., 2021). Even though I stated the audience size in each Study 5.2 discounting trial, participants may still have been biased towards assuming a greater number of other people would witness a situation simply because it was occurring online. An additional complication in studying whether audience size has a similar effect on cyberbystanders as it does on offline bystanders is that of asynchronicity (Best et al., 2014; Mittmann et al., 2022). Once a post containing aversive behaviour is made in an online space, there are those who see it in real-time but also those who see the post minutes, hours, and days (or longer) later – all of whom constitute the total audience. At this stage, my results suggest that there is no diffusion of responsibility effect in online spaces generated by larger audience sizes. Future research, where it is made explicit how many people witness the online aversive behaviour in real-time versus those who are ultimately witnesses (over time), may be worthwhile to check whether some form of audience effect is present online. Additionally, there may be an interaction between the effect that audience size and perceived severity have on cyberbystander behaviour, in that the latter is inflated when there is a larger audience, and this relationship should also be examined.

Severity Discounting

Discounting methods have been successfully applied to decisions made as a function of monetary (e.g., Friedel et al., 2014), probabilistic (e.g., Lawyer & Schoepflin, 2013; McKerchar et al., 2010), social (e.g., Hayashi & Tahmasbi, 2020; Locey et al., 2011), and several other types of outcomes (e.g., Garami & Moustafa, 2020). In the context of cyberbystander decision-making, I developed severity discounting as a new sub-category of discounting research. In Studies 4.1 and 5.1, I used a categorical independent variable consisting of a condition moderately low in social severity (ECL) compared to another that was moderately high in severity (ECH). In both, I found that participants were more likely to intervene in the high-severity condition, as I hypothesised and as others have found (Macháčková et al., 2015; Marciano et al., 2022). I included a severity rating activity in Study 5.2 as a test of the validity of my two-level severity manipulation. In the activity, I asked participants to order 10 short statements that outlined instances of online aversive behaviour. Two of the statements were shortened versions of the moderately-low (ECL) and moderately-high (ECH) scenarios. For the most part, participants rated the ECL statement as lower in severity than the ECH statement, which provides support for the external validity of my severity-related findings thus far.

To further examine the relationship between perceived scenario severity and cyberbystanders' intention to intervene, I extended the categorical variables from earlier studies (4.1 & 5.1) to one resembling a scale variable. The results from the severity rating activity in Study 5.2 served as the foundation for my novel independent variable, the Severity of Online Aversive Behaviour (SOAB) scale. I developed the SOAB as a continuum ranging from low to high severity, with seven distinct points that are exemplified by short scenarios. For example, the lowest point on the SOAB scale (i.e., the least severe) is exemplified by the statement, “Y never receives any response from X when they try talking to them during online multiplayer games”. Conversely, the highest point on the SOAB scale (i.e., the most severe) is exemplified by “X has filmed Y without their knowledge, then posted the video online to make fun of them”. As well as using the SOAB scale in the discounting section of the Study 6.1 questionnaire, I asked this group of participants (in addition to Study 5.2 participants) to order the shortened severity statements as a check that the SOAB scale statements were externally valid. That is, to check whether the Study 6.1 participants perceived the severity of the series of statements in a similar way to participants from Study 5.2. The findings from both studies aligned, supporting external validity in that two independent groups of participants rated the severity of the SOAB scale statements consistently.

The introduction of the SOAB scale addresses a gap in the existing cyberbystander research literature. Prior quantitative analyses of perceived severity have been limited, and among such studies, the utilisation of a scale-like independent variable to manipulate perceived severity, as opposed to categorical distinctions, appears to be unique (e.g., Bastiaensens et al., 2014; Hayashi & Tahmasbi, 2020). Using the SOAB scale as the primary independent variable, the likelihood of intervention showed a consistent pattern: intention to intervene was highest in scenarios of greater perceived severity and declined as the severity of the situation lessened. Thus, I found that the cyberbystander decisions about intervening were discounted as a function of perceived severity. Notably, as the perceived severity increased, participants rated their own propensity to intervene as greater than that of another person. Conversely, at lower levels of perceived scenario severity, participants indicated a similar likelihood of intervening as that of others (e.g., their personal and someone else’s responses were comparable). Therefore, the use of the SOAB scale offers a new and more comprehensive approach to examining the influence of severity on cyberbystander decision-making.

Additional Measures Related to Intervening

To complement the intention-to-intervene results from the discounting sections of my questionnaires, I included measures of several other variables which are commonplace in the extant bullying and bystander behaviour literature (e.g., DeSmet et al., 2012; Kowalski et al., 2014). Overall, my results provided either weak or mixed support for many of these. For example, I was not able to provide clarification to the currently open question (e.g., Barlińska et al., 2013; Erreygers et al., 2016; Van Cleemput et al., 2014) regarding the way in which (if at all) one's preferred gender identity interacts with the likelihood of carrying out cyberbystander intervention. I found negating evidence regarding whether an individual's sense of where they belong within a societal hierarchy (e.g., their perceived socioeconomic status; PSES) is related to cyberbystander behaviour – it was not.

Research has consistently shown a link between individuals who experience being the target of aversive behaviour, such as bullying, and those with lower self-esteem (e.g., Burger & Bachmann, 2021; Lei et al., 2020). Others' research has also shown a link between those with higher self-esteem (and, relatedly, higher self-efficacy) and the propensity to intervene in aid of targets of aversive behaviour (e.g., Caravita et al., 2009; Ferreira et al., 2020; Pöyhönen et al., 2012; Van Cleemput et al., 2014). In every discounting study, I included Rosenberg's (1965) measure of self-esteem (the RSE) and found no consistent significant relationship between participant self-esteem and their likelihood to intervene in aversive online behaviour. I found that RSE scores were inconsistently related to PSES (Adler et el., 2000), SAS (Kwon et al., 2013), and age but were only associated with time metrics (TO, ToSM, & PoT) within the Study 6.1 sample. My results are contrary to previous research relating to self-esteem and intervention; thus, they warrant further, larger-scale testing using cyberbystander discounting methods.

I found limited indication that age was related to the participant's intention to intervene, in contrast to other researchers (e.g., DeSmet et al., 2016; Schultze-Krumbholz et al., 2018). The exception was those who completed the Study 6.1 questionnaire for whom there was a negative relationship such that older participants were less likely to intervene (and less likely to think that others would). Study 6.1 participants, as a group, also showed a similarly strong (i.e., approximately 20-30%) negative correlation between their ages and ToSM, PoT, and their SAS scores. Relatedly, in Study 6.1, participant age was positively correlated, to approximately the same extent, with increased self-esteem scores (i.e., a positive correlation between age and RSE).

The three inter-related measures of time spent online (TO, ToSM, and PoT) tended to correlate highly, although participant's smartphone addiction scores (SAS; Kwon et al., 2013) were only related to the SNS-related metrics (ToSM and PoT). ToSM emerged as a promising predictor of intention to intervene in online aversive behaviour. ToSM and SAS (and PoT) were correlated in every instance (Studies 5.1, 5.2, & 6.1), which shows support for the use of the SAS measure with a broader range of populations, for example, with Aōtearoa New Zealand University students sampled in Studies 5.1 and 6.1. An overreliance on one's smartphone (so-called "smartphone addiction"), therefore, overlaps considerably with how much time is spent using SNS. Yet, that SAS was not significantly correlated with intentions to intervene suggests that the mere act of being in online social spaces (i.e., ToSM) has a greater influence on cyberbystander decision-making than does being highly involved in these spaces. Step 1 of the BIM regards the bystander first noticing the aversive behaviour – though not necessarily recognising it as aversive, which is part of Step 2 (Latané & Darley, 1970). Using questionnaires and discounting methods, I was not able to empirically test Step 1 in my studies. However, it may simply be that those who spend the most time on SNS are the most likely to notice online aversive behaviour in the first place (Park et al., 2014) and are, therefore, also more likely to progress to Step 2 and the later BIM steps (Latané & Darley, 1970).

Limitations

Some researchers have suggested that individuals who are more oriented to the point of view of the target are more likely to intervene to help compared to those who tend to be more concerned with how the aggressor may react to intervention (Cui et al., 2023; DeSmet et al., 2016). It is likely that those who have personally experienced being a target of online aversive behaviour orient towards targets when they witness aversive behaviour; thus, previous targets may show a higher intention to intervene for others. I did not ask participants in my discounting studies about their previous experiences (if any) regarding online aversive behaviour, but it is possible that their experiences could help explain the differences between those who do and those who do not intervene.

I also did not collect any information from participants regarding their cultural backgrounds. It may be that there is an instructive difference in the views about cyberbystander intervention between those, for example, from more individualistic cultures compared to those from more collectivist cultures (e.g., DeSmet et al., 2012; Huang & Chou, 2010; Lapidot-Leffler, 2017). A further step for researchers of cyberbystander behaviour would be to adopt a cross-cultural approach to determine any relevant factors that influence the decision to intervene and to test these using discounting methods (Du et al., 2002; Kirby et al., 2022).

Future Research

In my studies, perceived scenario severity was consistently an important factor in cyberbystander's decision making. Using various alternate methods, other researchers have also posited that severity is a critical factor for bystanders (e.g., Hayashi & Tahmasbi, 2020; Koehler & Weber, 2018). For this reason, I extended my method for examining perceived scenario severity from manipulating a categorical variable (i.e., an independent variable with two levels) to using a 7-point severity of online aversive behaviour scale (SOAB). As a new continuum of scenario severity developed to manipulate scenario severity within discounting research, the SOAB scale requires replication work. If others are to use the SOAB scale for similar purposes, it will be important to include a manipulation check of the sort in Study 6.1. In other words, those using the SOAB scale to manipulate scenario severity should also ask their participants to rate the severity of the individual statements that make up the scale.

In addition to the probability discounting (Chapters 4 & 5) and the core impact of scenario severity (Chapter 6), the personal effort required of bystanders to act warrants attention to better understand the contingencies related to cyberbystander behaviour. Given that Lueng et al. (2018) found that most people understood intervening to help a target to be the correct response, the effort required to intervene could be a barrier to cyberbystander intervention. Although I found inconsistent effects of my responsibility manipulations (i.e., "Would you intervene?" vs. "Would someone else intervene?"), it is possible that future research using this dichotomy could be helpful in developing an effort-discounting paradigm to determine whether increasing difficulty (i.e., requiring more effort) leads to steeper discount rates of intervening. Additionally, the effort needed to carry out intervention may be critical to whether cyberbystanders fulfil Step 3 of the BIM – something that could also be tested using the methods developed in this thesis. Even if effort is not relevant in cyberbystander decision-making, the results regarding whose responsibility it was to intervene were inconclusive and therefore warrant continued research attention, which may include reframing this variable as a difference between "would you?" versus "should you?". In my studies, participants rated themselves as more likely in two of the three studies, which indicates that the next study using the responsibility manipulation could centre on a unidirectional hypothesis rather than the unspecified one I used in Studies 5.1, 5.2, and 6.1.

A further avenue for research is in continuing to measure the fulfilment of each discrete step of the BIM (i.e., currently conceptualised as Steps 1-5). Several steps of the BIM are yet to be thoroughly examined (Mennicke et al., 2022) and baseline measures of how often bystander intervention occurs have not always been clearly reported (reviewed in Chapter 3). The process of estimating the fulfilment of Steps 2 and 3 in my series of studies (Studies 5.1, 5.2, and 6.1) could be extended to other steps, as well as applied to broader participant populations. Furthermore, measures of the fulfilment of each step in the model is necessary to conduct a rigorous analysis of the structure of the BIM. Such an analysis is long overdue for substantiating that the BIM is best-framed as five individual steps and that these steps are indeed stepwise.

Conclusion

Much research, including my own, supports the idea that most online communication is reinforcing and broadly beneficial rather than displacing other forms of social connection (e.g., Boer et al., 2021; Dienlin et al., 2017; Rosenfeld & Thomas, 2012). Nevertheless, harmful behaviour does occur online. Regardless of its true prevalence rate, online aversive behaviour is detrimental to the individuals and groups targeted by it and therefore warrants researchers' attention (Bastiaensens et al., 2014; Luo & Bussey, 2022; Macháčková & Pfetsch, 2016; Ng et al., 2022). Cyberbystanders intervening to help the targets of online aversive behaviour has been highlighted as important for mitigating harm (e.g., Luo & Bussey, 2022; Pyżalski et al., 2022; Sheeran et al., 2014) and for helping establish prosocial behaviour as more normative in online spaces (e.g., Crone et al., 2022; Rabb et al., 2022). Changing the intervention behaviour of cyberbystanders is a behavioural solution (rather than relying on algorithms or business interests) to the widespread and insidious issue of cyberbullying and similar online aversive behaviours.

The integration of cross-discipline concepts is crucial for understanding and addressing complex social issues (Kirby et al., 2022; Neuringer, 1991). I combined theories from social psychology and behaviour analysis to help in understanding cyberbystander behaviour. In particular, the novel application of discounting methods (e.g., probability and severity discounting) to cyberbystander decision-making was successful both for exploring the applicability of the BIM and for examining the relationship between various environmental factors and decision-making (Baum, 2017; Wolter & Napper, 2023).

My research suggests that initiatives for cyberbystander behaviour change should begin by addressing people's perceptions of severity, as these perceptions may be directly related to how likely someone is to intervene (e.g., Macaulay et al., 2019; Macháčková, 2020; Zhao et al., 2023). Koehler and Weber (2018) have suggested that educators in schools could help young people to more clearly assess the severity of cyberbullying and related aversive online behaviours; though such initiatives need not be limited to school environments. Similarly, Zhao et al. (2023) noted the need to educate people about the severity of cyberbullying and its impacts by providing examples of aversive behaviour and guiding them through a considered discussion of these. It may be that discussing personal or their peers' experiences with aversive online behaviour (rather than discussing generic examples) increases the efficacy of this kind of programme because the scenarios would be directly relatable (McLoughlin et al., 2020; Midgett et al., 2017; Pyżalski et al., 2022). The focus groups that I facilitated as an exploratory first step for this thesis may, in fact, be a useful template for behaviour change initiatives. Behaviour change initiatives such as these address Steps 1 and 2 of the BIM: noticing and then recognising a situation as requiring intervention (Latané & Darley, 1970). A promising avenue for future research is to leverage the discounting methods I used to examine what proportion of participants fulfilled the early steps in the BIM to determine whether such an initiative is effective. That is, researchers could measure step-fulfilment before and after people participate in a facilitated discussion about online aversive behaviour. More generally, my thesis makes clear that further concerted research in this area (which could utilise probability discounting) is essential to develop effective strategies that foster a culture of active bystander intervention to reduce the prevalence of aversive behaviour online.

References

Abbate, C.S., Misuraca, R., Vaccaro, C., Roccella, M., Vetri, L., & Miceli, S. (2022). Prosocial priming and bystander effect in an online context. *Frontiers in Psychology, 13*, 945630. https://doi.org/10.3389/fpsyg.2022.945630

Adachi, P.J.C., & Willoughby, T. (2013). More than just fun and games: The longitudinal relationships between strategic video games, self-reported problem solving skills, and academic grades. *Journal of Youth and Adolescence, 42*, 1041-1052. https://doi.org/10.1007/s10964-013-9913-9

Adams, T.L., Li, Y., & Liu, H. (2020). A Replication of beyond the turk: Alternative platforms for crowdsourcing behavioral research–sometimes preferable to student groups. *AIS Transactions on Replication Research, 6,* Article 15. https://doi.org/10.17705/1atrr.00058

Adler, N.E., Epel, E.S., Castellazzo, G., & Ickovics, J.R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology, 19*, 586-592. https://doi.org/10.1037/0278-6133.19.6.586

Alhassan, A.A., Alqadhib, E.M., Taha, N.W., Alahmari, R.A., Salam, M., & Almutairi, A.F. (2018). The relationship between addiction to smartphone usage and depression among adults: A cross sectional study. *BMC Psychiatry, 18,* 148. https://doi.org/10.1186/s12888-018-1745-4

Alicke, M.D., Klotz, M.L., Breitenbecher, D.L., Yurak, T.J., & Vredenburg, D.S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*, 804–825. https://doi.org/10.1037/0022-3514.68.5.804

Allison, K.R. & Bussey, K. (2016). Cyber-bystanding in context: A review of the literature on witnesses' responses to cyberbullying. *Children and Youth Services Review, 65,* 183-194. https://doi.org/10.1016/j.childyouth.2016.03.026

Allison, K.R., & Bussey, K. (2017). Individual and collective moral influences on intervention in cyberbullying. *Computers in Human Behavior, 74,* 7-15. https://doi.org/10.1016/j.chb.2017.04.019

Allport, F.H. (1924). The group fallacy in relation to social science. *American Journal of Sociology, 29,* 688-703. https://doi.org/10.1086/213647

Anderson, J., Bresnahan, M., & Musatics, C. (2014). Combating weight-based cyberbullying on Facebook with the dissenter effect. *Cyberpsychology, Behavior, and Social Networking, 17,* 281-286. https://doi.org/10.1089/cyber.2013.0370

Anderson, K.J. (2001). Internet use among college students: An exploratory study. *Journal of American College Health, 1,* 21-26. https://doi.org/10.1080/07448480109595707

Andrade, L.F., & Petry, N.M. (2012). Delay and probability discounting in pathological gamblers with and without a history of substance use problems. *Psychopharmacology, 219,* 491-499. https://doi.org/10.1007/s00213-011-2508-9

Andrews, S., Ellis, D.A., Shaw, H., Piwek, L. (2015) Beyond self-report: tools to compare estimated and real-world smartphone use. *PLoS ONE 10,* e0139004. https://doi.org/10.1371/journal.pone.0139004

Andrews, T.E.A.-M. (2022). *Determinants of risky decision-making: What is peer influence?* [Doctoral dissertation, University of Waikato]. Research Commons.

Antoniadou, N., Kokkinos, C.M. & Fanti, K.A. (2019). Traditional and cyber bullying/victimization among adolescents: Examining their psychosocial profile through latent profile analysis. *International Journal of Bullying Prevention, 1,* 85–98. https://doi.org/10.1007/s42380-019-00010-0

Arnold, M.A. (2012). *Delay discounting in light-social and heavy-social drinkers* [Honors Dissertation]. James Madison University Scholarly Commons.

Bailey, A.J., Romeu, R.J., & Finn, P.R. (2021). The problems with delay discounting: a critical review of current practices and clinical applications. *Psychological Medicine, 51,* 1799-1806. https://doi.org/10.1017/S0033291721002282

Baldassarri, D., & Grossman, G. (2013). The effect of group attachment and social position on prosocial behavior: Evidence from lab-in-the-field experiments. *PLOS ONE, 8,* e58750. https://doi.org/10.1371/journal.pone.0058750

Ballard, T., Luckman, A., & Konstantinidis, E. (2023). A systematic investigation into the reliability of inter-temporal choice model parameters. *Psychonomic Bulletin & Review.* https://doi.org/10.3758/s13423-022-02241-7

Banchik, A.V. (2020). Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights–related content. *New Media & Society, 23*, 1527-1544. https://doi.org/10.1177/1461444820912724

Bandura, A. (1982). The assessment and predictive generality of self-perception of efficacy. *Journal of Behavioral Theory and Experimental Psychiatry, 13,* 195-199. https://doi.org/10.1016/0005-7916(82)900004-0

Banyard, V.L. (2008). Measurement and correlates of prosocial bystander behavior: The case of interpersonal violence. *Violence and Victims, 23,* 83-97. https://doi.org/10.1891/0886-9708.23.1.83

Barcelos Nomicos, L., Jacobs, K.W., & Locey, M. (2020). The effects of obligatory and preferential frames on delay discounting. *The Analysis of Verbal Behavior, 36,* 74-86. https://doi.org/10.1007/s40616-020-00127-0

Barlett, C.P., Madison, C.S., Heath, J.B., & DeWitt, C.C. (2019). Please browse responsibly: A correlational examination of technology access and time spent online in the Barlett

Gentile Cyberbullying Model. *Computers in Human Behavior, 92,* 250-255. https://doi.org/10.1016/j.chb.2018.11.013

Barlett, C.P., Seyfert, L. W., Simmers, M. M., Hsueh Hua Chen, V., Cavalcanti, J. G., Krahé, B., ... & Skowronski, M. (2021). Cross‐cultural similarities and differences in the theoretical predictors of cyberbullying perpetration: Results from a seven‐country study. *Aggressive behavior, 47,* 111-119. https://doi.org/10.1002/ab.21923

Barlińska, J., Szuster, A., & Winiewski, M. (2013). Cyberbullying among adolescent bystanders: Role of the communication medium, form of violence, and empathy. *Journal of Community & Applied Social Psychology, 23,* 37-51. https://doi.org/10.1002/casp.2137

Barlińska, J., Szuster, A., & Winiewski, M. (2015). The role of short- and long-term cognitive empathy activation in preventing cyberbystander reinforcing cyberbullying behavior. *Cyberpsychology, Behavior, and Social Networking, 18,* 241-244. https://doi.org/10.1089/cyber.2014.0412

Barnett, J.E.H., Fisher, K.W., O'Connell, N., & Franco, K. (2019). Promoting upstander behavior to address bullying in schools. *Middle School Journal, 50,* 6-11. https://doi.org/10.1080/00940771.2018.1550377

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research, 1,* 19. Retrieved from http://www.brandeis.edu/pubs/jove/HTML/v1/bartle.html

Bartle, R. (2013). Design principles: Use and misuse. In T. Quandtm, & S. Kröger (Eds.), *Multiplayer: The Social Aspects of Digital Gaming* (1st ed., pp. 10-22). Routledge.

Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2016). From normative influence to social pressure: How relevant others affect whether bystanders join in cyberbullying. *Social Development, 25*, 193-211. https://doi.org/10.1111/sode.12134

Bastiaensens, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2014). Cyberbullying on social network sites. An experimental study into bystanders' behavioural intentions to help the victim or reinforce the bully. *Computers in Human Behavior, 31*, 259-271. ttps://doi.org/10.1016/j.chb.2013.10.036

Baum, W.M. (2017). *Understanding Behaviorism: Behavior, culture, and evolution* (3rd Ed.). Wiley Blackwell.

Bauman, S., & Bellmore, A. (2015). New directions in cyberbullying research. *Journal of School Violence, 15,* 1-10. https://doi.org/10.1080/15388220.2014.968281

Belisle, J., Campbell, L. & Todd, M. (2020). Delaying the point of no return: Discounting evident in climate change policy preference. *Behavior and Social Issues, 29*, 64–77. https://doi.org/10.1007/s42822-020-00035-x

Bem, D.J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183-200. https://doi.org/10.1037/h0024835

Bem, D.J. (1972). Self-perception theory. *Advances in Experimental Social Psychology, 6*, 1-62. https://doi.org/10.1016/S0065-2601(08)60024-6

Best, P., Manktelow, R., Taylor, B. (2014). Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review, 41,* 27-36. https://doi.org/10.1016/j.childyouth.2014.03.001

Bhattacherjee, A., & Sanford, C. (2009). The intention-behaviour gap in technology usage: The moderating role of attitude strength. *Behaviour & Information Technology, 28,* 389-401. https://doi.org/10.1080/01449290802121230

Bickel, W.K., & Marsch, L.A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction, 96,* 73-86. https://doi.org/10.1046/j.1360-0443.2001.961736.x

Bickel, W.K., Odum, A.L., & Madden, G.J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology, 146,* 447-454. https://doi.org/10.1007/PL00005490

Biglan, A., & Lichtenstein, E. (1984). A Behavior-Analytic Approach to Smoking Acquisition: Some Recent Findings. *Journal of Applied Social Psychology, 14,* 207–223. https://doi.org/10.1111/j.1559-1816.1984.tb02232.x

Blackwell, C.K., Lauricella, A.R., Conway, A., & Wartella, E. (2014). Children and the internet: Developmental implications of web site preferences among 8- to 12-year-old children. *Journal of Broadcasting & Electronic Media, 58,* 1-20. https://doi.org/10.1080/08838151.2013.875022

Blair, C.A., Thompson, L.F., & Wuensch, K.L. (2005). Electronic helping behavior: The virtual presence of others makes a difference. *Basic and Applied Social Psychology, 27,* 171–178. https://doi.org/10.1207/s15324834basp2702_8

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 1,* 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bleakley, A., Ellithorpe, M., & Romer, D. (2016). The role of parents in problematic internet use among US adolescents. *Media and Communication, 4,* 24-34. https://doi.org/10.17645.mac.v4i3.523

Boer, M., Stevens, G.W.J.M., Finkenauer, C., de Looze, M.E., & van den Eijnden, R.J.J.M. (2021). Social media use intensity, social media use problems, and mental health among adolescents: investigating directionality and mediating processes. *Computers in Human Behavior, 116,* 106645. https://doi.org/10.1016/j.chb.2020.106645

Borges, A.M., Kuang, J., Milhorn, H., & Yi, R. (2016). An alternative approach to calculating Area-Under-the-Curve (AUC) in delay discounting research. *Journal of the Experimental Analysis of Behavior, 106,* 145-155. https://doi.org/10.1002/jeab.219

Brinberg, M., Ram, N., Yang, X., Cho, M.-J., Sundar, S.S., Robinson, T.N., & Reeves, B. (2021). The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior, 114,* 106570. https://doi.org/10.1016/j.chb.2020.106570

Brockner, J., & Hulton, A.J.B. (1978). How to reverse the vicious cycle of low self-esteem: The importance of situational focus. *Journal of Experimental Social Psychology, 14,* 564-578. https://doi.org/10.1016/0022-1031(78)90050-1

Brown, J.D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin, 38,* 143-280. https://doi.org/10.1177/0146167211432763

Brühlmann, F., Petralito, S., Aeschbach, L.F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology, 2,* 100022. https://doi.org/10.1016/j.metip.2020.100022

Brumfitt, S., & Sheeran, P. (1999). Development and validation of the visual analogue self-esteem scale (VASES). *British Journal of Clinical Psychology, 38,* 387-400. https://doi.org/10.1037//0278-6133.19.3.283

Buchner, A., Erdfelder, E., Faul, F., & Lang, A.-G. (n.d.). G*Power. *Heine University Düsseldorf*. https://tinyurl.com/424kbpnz

Buckley, L., Fowler, J., & Bankole, E. (2022). The nature and context of online communication in promoting prosocial behaviour in youth: A systematic review. *SocArXiv*. https://doi.org/10.31235/osf.io/nqvcp

Bunce, C. (2009). Correlation, agreement, and Bland-Altman Analysis: statistical analysis of method comparison studies, *American Journal of Ophthalmology, 148,* 4-6. https://doi.org/10.1016/j.ajo.2008.09.032

Burger, C., & Bachmann, L. (2021). Perpetration and victimization in offline and cyber contexts: A variable- and person-oriented examination of associations and differences regarding domain-specific self-esteem and school adjustment. *International Journal of Environmental Research and Public Health, 18*, 10429. https://doi.org/10.3390/ijerph181910429

Burke, M., Kraut, R., & Marlow, C. (2011). Social capital on Facebook: Differentiating uses and users. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 571-580). https://doi.org/10.1145/1978942.1979023

Burn, S.M. (2009). A Situational Model of Sexual Assault Prevention through Bystander Intervention. *Sex Roles 60,* 779–792. https://doi.org/10.1007/s11199-008-9581-5

Butler, L. C., Graham, A., Fisher, B. S., Henson, B., & Reyns, B. W. (2022). Examining the effect of perceived responsibility on online bystander intervention, target hardening, and inaction. *Journal of Interpersonal Violence, 37,* 21–22, NP20847–NP20872. https://doi.org/10.1177/08862605211055088

Byrne, H., Dooley, B., Fitzgerald, A., & Dolphin, L. (2016). Adolescents' definitions of bullying: the contribution of age, gender, and experience of bullying. *European Journal of Psychology of Education, 31*, 403-418. http://dx.doi.org/10.1007/s10212-015-0271-8

Cabello-Hutt, T., Cabello, P., & Claro, M. (2017). Online opportunities and risks for children and adolescents: The role of digital skills, age, gender and parental mediation in Brazil. *New Media & Society, 20,* 2411-2431. https://doi.org/10.1177/1461444817724168

Callender, C. (2022). Is discounting for tense rational? In C. Hoerl, T. McCormack, & A. Fernandes (Eds.), *Temporal asymmetries in philosophy and psychology* (pp. 114-138). Oxford university Press. https://doi.org/10.1093/oso/9780198862901.003.0005

Campos-Mercade, P. (2021). The volunteer's dilemma explains the bystander effect. *Journal of Economic Behavior & Organization, 186,* 646–661. https://doi.org/10.1016/j.jebo.2020.11.012

Caravita, S.C.S., Pöyhönen, V., & Salmivalli, C. (2009). Prosocial involvement in antisocial situations: Defending the victims of bullying. In T.A. Kinney & M. Pörhölä (Eds.), *Anti- and pro-social communication: Theories, methods, and applications* (pp. 173-182). Peter Lang.

Carson, R.T., & Roth Tran, B. (2009). Discounting behavior and environmental decisions. *Journal of Neuroscience, Psychology, and Economics, 2,* 112–130. https://doi.org/10.1037/a0017685

Cassidy, W., Faucher, C., & Jackson, M. (2013). Cyberbullying among youth: A comprehensive review of current international research and its implications and application to policy and practice. *School Psychology International, 34,* 575-612. https://doi.org/10.1177/0143034313479697

Castells, M. (2002). *The Internet galaxy: Reflections on the Internet, business, and society*. Oxford University Press, USA.

Catone, G., Senese, V.P., Pisano, S., Siciliano, M., Russo, K., Muratori, P., Marotta, R., Pascotto, A., & Broome, M.R. (2020). The drawbacks of information and communication technologies: Interplay and psychopathological risk of nomophobia and cyber-bullying, results from the bullying and youth mental health Naples study (BYMHNS). *Computers in Human Behavior, 113,* 106496. https://doi.org/https://doi.org/10.1016/j.chb.2020.106496

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46,* 112-130. https://doi.org/10.3758/s13428-013-0365-7

Charness, G., Gneezy, U., & Kuhn, M.A. (2012). Experimental methods: between-subject and within-subject design. *Journal of Economic Behavior and Organization, 81,* 1-8. https://doi.org/10.1016/j.jebo.2011.08.009

Chen, J., Mortensen, K., & Bloodworth, R. (2014). Exploring contextual factors and patient activation. *Health Education & Behavior, 41,* 614-624. https://doi.org/10.1177/1090198114531781

Chen, Y., Huo, Y., & Liu, J. (2022). Impact of online anonymity on aggression in ostracized grandiose and vulnerable narcissists. *Personality and Individual Differences, 188,* a111448. https://doi.org/10.1016/j.paid.2021.111448

Cheng, S., (Kuo-Hsun) Ma, J., & Missari, S. (2014). The effects of Internet use on adolescents' first romantic and sexual relationships in Taiwan. *International Sociology, 29,* 324–347. https://doi.org/10.1177/0268580914538084

Chmielewski, M., & Kucker, S.C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11,* 464-473. https://doi.org/10.1177/1948550619875149

Chun, J., Lee, J., Kim, J., & Lee, S. (2020). An international systematic review of cyberbullying measurements. *Computers in Human Behavior, 113,* a106485. https://doi.org/10.1016/j.chb.2020.106485

Chung, W., Sun, C.-K., Tsai, I.-T., Hung, K.-C., Chiu, H.-J., Tzang, R.-F., Yeh, P.-Y., & Cheng, Y.-S. (2021). A systematic review and meta-analysis on the clinical implications of probability discounting among individuals with Internet gaming disorder. *Scientific Reports, 11,* 3177–3177. https://doi.org/10.1038/s41598-021-82822-z

Codol, J.P. (1975). On the so-called superiority conformity of the self behavior: Twenty experimental investigations. *European Journal of Social Psychology, 5,* 457-501. https://doi.org/10.1002.ejsp.2420050404

Coopersmith, S. (1967). *The antecedents of self-esteem.* Freeman.

Coyne, S.M., Rogers, A.A., Zurcher, J.D., Stockdale, L., & Booth, Booth, MC. (2020). Does time spent using social mdia impact mental health?: An eight year longitudinal study. *Computers in Human Behavior, 104,* 106160. https://doi.org/10.1016/j.chb.2019.106160

Craft, W.H., Tegge, A.N., & Bickel, W.K. (2022a). Narrative theory: Within-subject effects of active and control scarcity narratives on delay discounting in alcohol use disorder. *Experimental and Clinical Psychopharmacology, 30*, 500-506. https://doi.org/10.1037/pha0000478

Craft, W.H., Tegge, A.N., Freitas-Lemos, R., Tomlinson, D.C., & Bickel, W.K. (2022b). Are poor quality data just random responses?: A crowdsourced study of delay discounting in alcohol use disorder. *Experimental and Clinical Psychopharmacology, 30,* 409-414. https://doi.org/10.1037/pha0000549

Craig, W. M., & Pepler, D. J. (1998). Observations of bullying and victimization in the school yard. *Canadian Journal of School Psychology, 13,* 41–59. https://doi.org/10.1177/082957359801300205

Cramer, D. (1990). Helpful actions of close friends to personal problems and distress. *British Journal of Guidance & Counselling, 18,* 281-293. https://doi.org/10.1080/03069889008253580

Critchfield, T.S., & Kollins, S.H. (2001). Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis, 34,* 101-102. https://doi.org/10.1901/jaba.2001.34-101

Crone, E.A., Sweijen, S.W., Te Brinke, L.W., & van de Groep, S. (2022). Pathways for engaging in prosocial behavior in adolescence. *Advances in child development and behavior, 63*, 149–190. https://doi.org/10.1016/bs.acdb.2022.03.003

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology, 24,* 349-354. https://doi.org/10.1037/h0047358

Cui, K., Xie, H. & Peng, H. (2023). How past cyber victimization affects bystanders' position taking in offline bullying situations: A moderated mediation model of self-concept and social support. *School Mental Health 15,* 416–430. https://doi.org/10.1007/s12310-022-09557-8

Cunningham, S., Hudson, C.C., & Harkness, K. (2021). Social Media and Depression Symptoms: A Meta-Analysis. *Research on Child and Adolescent Psychopathology, 29,* 241-253. https://doi.org/10.1007/s10802-020-00715-7

Darley, J.M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8,* 377-383. https://doi.org/10.1037/h0025589

Demirci, K., Akgönül, M., & Akpinar, A. (2015). Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. *Journal of Behavioral Addictions, 4,* 85-92. https://doi.org/10.1556/2006.4.2015.010

Demirci, K., Orhan, H., Demirdas, A., Akpinar, A., & Sert, H. (2014). Validity and reliability of the Turkish Version of the Smartphone Addiction Scale in a younger population. *Bulletin of Clinical Psychopharmacology, 24,* 226-234. https://doi.org/10.21203/rs.3.rs-676215/v1

DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., & De Bourdeaudhuij, I. (2012). Mobilizing bystanders of cyberbullying: An exploratory study into behavioural determinants of defending the victim. In B. K. Wiederhold & G. Riva (Eds.*), Annual Review of Cybertherapy and Telemedicine 2012: Advanced technologies in the behavioral, social and neurosciences* (p. 58–63). Amsterdam: IOS Press. https://doi.org/10.3233/98-1-61499-121-2-58

DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., Cardon, G., & De Bourdeaudhuij, I. (2016). Deciding whether to look after them, to like it, or leave it: A multidimensional analysis of predictors of positive and negative bystander behavior in cyberbullying among adolescents. *Computers in Human Behavior, 57,* 398-415. https://doi.org/10.1016/j.chb.2015.12.051

Deutsch, M., & Gerard, H.B. (1955). A study of normative and informational social influences upon individual judgement. *The Journal of Abnormal and Social Psychology, 51,* 629-636. https://doi.org/10.1037/h0046408

Dienlin, T., Masur, P.K., & Trepte, S. (2017). Reinforcement or displacement? The reciprocity of FtF, IM, and SNS communication and their effects on loneliness and life satisfaction. *Journal of Computer Mediated Communication, 22,* 71–87. https://doi.org/10.1111/jcc4.12183

Dillon, K.P., & Bushman, B.J. (2015). Unresponsive or un-noticed?: Cyberbystander intervention in an experimental cyberbullying context. *Computers in Human Behavior, 45,* 144-150. https://doi.org/10.1016/j.chb.2014.12.009

Dixon, M.R., Jacobs, E.A., & Sanders, S. (2006). Contextual control of delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis, 39,* 413-422. https://doi.org/10.1901/jaba.2006.173-05

Domahidi, E. (2018). The associations between online media use and users' perceived social resources: A meta-analysis. *Journal of Computer-Mediated Communication, 23,* 181-200. https://doi.org/10.1093/jcmc/zmy007

Domínguez-Hernández, F., Bonell, L., & Martínez-González, A. (2018). A systematic literature review of factors that moderate bystanders' actions in cyberbullying. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace,* 12, Article 1. https://doi.org/10.5817/CP2018-4-1

Du, W., Green, L. & Myerson, J. (2002). Cross-Cultural Comparisons of Discounting Delayed and Probabilistic Rewards. *The Psychological Record, 52,* 479–492. https://doi.org/10.1007/BF03395199

Dubois, D., Rucker, D.D., & Galinsky, A.D. (2015). Social class, power, and selfishness: When and why upper and lower class individuals behave unethically. *Journal of Personality and Social Psychology, 108,* 436–449. https://doi.org/10.1037/pspi0000008

Dyer, H.T., & Abidin, C. (2022). Understanding Identity and Platform Cultures. In *The Sage Handbook of Digital Society* (2nd ed.). Sage Publications.

Dymond, S., & Roche, B. (2009). A contemporary behavior analysis of anxiety and avoidance. *The Behavior Analyst, 32*, 7-27. https://doi.org/10.1007/BF03392173

Dymond, S., Schlund, M.W., Roche, B., & Whelan, R. (2014). The spread of fear: Symbolic generalization mediates graded threat-avoidance in specific phobia. *The Quarterly Journal of Experimental Psychology, 67,* 247-259. http://dx.doi.org/10.1080/17470218.2013.800124

Eagly, A.H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100,* 283–308. https://doi.org/10.1037/0033-2909.100.3.283

Easton, S., Morton, K., Tappy, Z., Francis, D., & Dennison, L. (2018). Young people's experiences of viewing the Fitspiration media trend: Qualitative study. *Journal of Medical Internet Research, 20, e219*. https://doi.org/10.2196/jmir.9156.

Ellis, D.A., Davidson, B.I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior? *International Journal of Human-Computer Studies, 130,* 86-92. https://doi.org/10.1016/j.ijhcs.2019.05.004

Erreygers, S., Pabian, S., Vandebosch, H., & Baillien, E. (2016). Helping behavior among adolescent bystanders of cyberbullying: The role of impulsivity. *Learning and Individual Differences, 48,* 61-67. https://doi.org/10.1016/j.lindif.2016.03.003

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39,* 175–191. https://doi.org/10.3758/bf03193146

Ferreira, P.C., Simão, A., Paiva, A., Ferreira, A.I. (2020). Responsive bystander behaviour in cyberbullying: a path through self-efficacy. *Behaviour and Information Technology, 39,* 511-524. https://doi.org/10.1080/0144929X.2019.1602671

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). Sage.

Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology, 36,* 267-278. https://doi.org/10.1002/ejsp.297

Fischer, P., Krueger, J.I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M., & Kainbacher, M. (2011). The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin, 137,* 517–537. https://doi.org/10.1037/a0023304

Fisher, G.G., Matthews, R.A., & Gibbons, A.M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology, 21*, 3–23. https://doi.org/10.1037/a0039139

Fisher, I. (1930). *The theory of interest.* New York. Macmillan.

Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of personality and social psychology, 46*, 404. https://psycnet.apa.org/doi/10.1037/0022-3514.46.2.404

Foreman, A.M., Hayashi, Y., Friedel, J.E., & Wirth, O. (2019). Social distance and texting while driving: A behavioral economic analysis of social discounting. *Traffic Injury Prevention, 20,* 702-707. https://doi.org/10.1080/15389588.2019.1636233

Freeman, K.A., & Morris, T.L. (1999). Explaining the Effects of Child Sexual Abuse: A Behavior Analytic Conceptualization. *Journal of Child Sexual Abuse, 7*, 3–21. https://doi.org/10.1300/J070v07n04_02

Friedel, J.E., DeHart, W.B., Madden, G.J., & Odum, A.L. (2014). Impulsivity and cigarette smoking: Discounting of monetary and consumer outcomes in current and non-smokers. *Psychopharmacology, 231,* 4517-4526. https://doi.org/10.1007/s00213-014-3597-z

Garami, J., & Moustafa, A.A. (2020). Delay, probability, and effort discounting in drug addiction. In A.A. Moustafa (Ed.), *Cognitive, clinical, and neural aspects of drug addiction* (pp. 61-83). Academic Press. https://doi.org/10.1016/B978-0-12-816979-7.00004-2

Gilroy, S.P., & Hantula, D.A. (2018). Discounting model selection with area-based measures: A case for numerical integration. *Journal of the Experimental Analysis of Behavior, 109,* 433-449. https:// doi.org/10.1002/jeab.318

Gilroy, S.P., Franck, C.T., & Hantula, D.A. (2017). The discounting model selector: Statistical software for delay discounting applications. *Journal of the Experimental Analysis of Behavior, 107,* 388-401. https://doi.org/10.1002/jeab.257

Gilroy, S.P., Strickland, J.C., Naude, G.P., Johnson, M.W., Amlung, M., & Reed, D.D. (2022). Beyond systematic and unsystematic responding: Latent class mixture models to characterize response patterns in discounting research. *Frontiers in Behavioral Neuroscience, 16,* 806944. https://doi.org/10.3389/fnbeh.2022.806944

Gopal, L. (2021). Facebook's Oversight Board & the Rule of Law: The Importance of Being Earnest. *Business Law Today.* https://businesslawtoday.org/2021/10/facebooks-oversight-board-the-rule-of-law-the-importance-of-being-earnest/?utm_medium=social#_ftn1

Gore, P.A., Leuwerke, W.C., & Turley, S.E. (2005). A psychometric study of the college self-efficacy inventory. *Journal of College Student Retention: Research, Theory & Practice, 7,* 227–244. https://doi.org/10.2190/5CQF-F3P4-2QAC-GNVJ

Green, L., & Myerson, J. (1996). Exponential versus hyperbolic discounting of delayed outcomes: Risk and waiting time. *Integrative & Comparative Biology, 36,* 496-505. https://doi.org/10.1093/icb/36.4.496

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 418-427. https://doi.org/10.1037/0278-7393.25.2.418

Greenhow, C., & Burton, L. (2011). Help from my “friends”: Social capital in the social network sites of low-income students. *Journal of Educational Computing Research, 45,* 223–245. https://doi.org/10.2190/EC.45.2.f

Greenwald, A. (1976). Within-subjects designs: to use or not to use. *Psychological Bulletin, 83,* 314-320. https://doi.org/10.1037/0033-2909.83.2.314

Gross, E.F., Juvonen, J., & Gable, S.L. (2002). Internet use and well-being in adolescence. *Journal of Social Issues, 58,* 75-90. https://doi.org/10.1111/1540-4560.00249

Grüne-Yanoff, T. (2015). Models of temporal discounting 1937-2000: an interdisciplinary exchange between economics and psychology. *Science in Context, 28,* 675-713. https://doi.org/10.1017/S0269889715000307

Guerin, B. (1994). *Analyzing social behavior: Behavior analysis and the social sciences.* Context Press.

Guerin, B. (2016). *A practical guide to contextual social analysis: The contextual understanding of human behavior*. Routledge.

Guerin, B. (2020). *Turning psychology into a social science* (1st ed.). Routledge. https://doi.org/10.4324/9781003021278

Hagey, K., & Horwitz, J. (2021). Facebook tried to make its platform a healthier place. It got angrier instead. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215

Harris, B., Regan, T., Schueler, J., & Fields, S.A. (2020). Problematic mobile phone and smartphone use scales: a systematic review. *Frontiers in Psychology, 11,* 672. https://doi.org/10.3389/fpsyg.2020.00672

Hartgerink, C.H., Van Beest, I., Wicherts, J.M., & Williams, K.D. (2015). The ordinal effects of ostracism: A meta-analysis of 120 Cyberball studies. *PloS one, 10,* e0127002. https://doi.org/10.1371/journal.pone.0127002

Hawkins, D.L., Pepler, D.J., & Craig, W.M. (2001). Naturalistic observations of peer interventions in bullying. *Social development, 10,* 512-527. https://doi.org/10.1111/1467-9507.00178

Hayashi, Y., & Glodowski, K.R. (2023). Males (but not females) who use social networking sites excessively more more impulsive decisions that those who do not: role of gender and delay discounting. *Behaviour & Information Technology, 42,* 1-12. https://doi.org/10.1080/0144929X.2021.2015441

Hayashi, Y., & Tahmasbi, N. (2020). Decision-making process underlying bystanders' helping cyberbullying victims: A behavioral economic analysis of role of social discounting. *Computers in Human Behavior, 104*, e106157. https://doi.org/10.1016/j.chb.2019.106157

Hayashi, Y., Miller, K., Foreman, A.M., & Wirth, O. (2016). A behavioral economic analysis of texting while driving: delay discounting process. *Accident; Analysis & Prevention, 97,* 132-140. https://doi.org/10.1016/j.aap.2016.08.028

Hayashi, Y., Tahmasbi, N., Romanowich, P., & Hantula, D.A. (2023). Social and delay discounting in a behavioral economic analysis of bystanders' helping cyberbullying victims: The moderating role of gender. *Computers in Human Behavior, 145,* e107783. https://doi.org/10.1016/j.chb.2023.107783

Hayes, M.H. (1921). Experimental development of the graphic rating method. *Psychological Bulletin, 18, 98*-99. https://doi.org/10.1037/h0064147

Haynes, S.N., & O'Brien, W.D. (2000). *Principles and practice of behavioral assessment*. Applied Clinical Psychology. https://doi.org/10.1007/978-0-306-47469-9

Heffer, T., Good, M., & Willoughby, T. (2019). The longitudinal association between social-media use and depressive symptoms among adolescents and young adults: an empirical reply to Twenge et al. (2018). *Clinical Psychological Science, 7,* 462-470. https://doi.org/10.1177/2167702618812727

Heflin, H., Shewmaker, J., & Nguyen, J. (2017). Impact of mobile technology on student attitudes, engagement, and learning. *Computers & education, 107,* 91-99. https://doi.org/10.1016/j.compedu.2017.01.006

Hellström, L., & Lundberg, A. (2020). Understanding bullying from young people's perspectives: An exploratory study. *Educational Research, 62,* 414-433. https://doi.org/10.1080/00131881.2020.1821388

Ho, T.T.Q., & Gu, C. (2023). Cyberbullying victimization and depression: self-esteem as a mediator and approach coping strategies as moderators. *Journal of American College Health, 71,* 94-101. https://doi.org/10.1080/07448481.2021.1877718

Hodge, E.-L., Robertson, N., & Sargisson, R. J. (2017). Mobile technologies in schools: The student voice. *Teachers and Curriculum, 17.* https://doi.org/10.15663/tandc.v17i2.169

Hodge, E.-L., & Sargisson, R.J. (2022). The prevalence and targets of cyberbullying behaviours: An observational approach. *International Journal of Psychology and Behavior Analysis*, *8*, 194. https://doi.org/10.15344/2455-3867/2022/194

Hogg, M.A., & Reid, S.A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory, 16*, 7-30. https://doi.org/10.1111/j.1468-2885.2006.00003.x

Holt, D.D., Green, L., & Myerson, J. (2003). Is discounting impulsive?: Evidence from temporal and probability discounting in gambling and non-gambling college students. *Behavioural Processes, 64,* 355-367. https://doi.org/10.1016/S0376-6357(03)00141-4

Howitt, D. (2019). *Introduction to qualitative research methods in psychology.* Pearson UK.

Hoxmeier, J.C., Acock, A.C., & Flay, B.R. (2020). Students as prosocial bystanders to sexual assasult: Demographic correlates of intervention norms, intentions, and missed opportunities. *Journal of Interpersonal Violence, 35*, 731-754. https://doi.org/10.1177/0886260517689888

Huang, G., & Li, K. (2016). The effect of anonymity on conformity to group norms in online contexts: A meta-analysis. *International Journal of Communication, 10*, 398-415. https://ijoc.org/index.php/ijoc/article/view/4037

Huang, Y-Y. & Chou, C. (2010). An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Computers in Human Behavior, 26,* 1581-1590. https://doi.org/10.1016/j.chb.2010.06.005

Hussain, I., Shu, R., Tangirala, S., & Ekkirala, S. (2019). The voice bystander effect: how information redundancy inhibits employee voice. *Academy of Management Journal, 62*, 828–849. https://doi.org/10.5465/amj.2017.0245

Janz, N.K., & Becker, M.H. (1984). The health belief model: A decade later. *Health Education Quarterly, 11,* 1-47. https://doi.org/10.1177/109019818401100101

Jarmolowicz, D.P., Bickel, W.K., Carter, A.E., Franck, C.T., & Mueller, E.T. (2012). Using crowdsourcing to examine relations between delay and probability discounting. *Behavioural Processes*, *91*, 308-312. https://doi.org/10.1016/j.beproc.2012.09.001

Jarmolowicz, D.P., Reed, D.D, Francisco, A.J., Bruce, J.M., Lemley, S.M., & Bruce, A.S. (2018). Modeling effects of risk and social distance on vaccination choice. *Journal of the Experimental Analysis of Behavior, 110,* 39-53. https://doi.org/10.1002/jeab.438

Jenkins, N.W., Parrish, J.M., Hrynewycz, N.M., Brundage, T.S., & Singh, K. (2020). Longitudinal evaluation of patient-reported outcomes measurement information system for back and leg pain in minimally invasive transforaminal lumbar interbody fusion. *Neurospin, 17,* 862-870. https://doi.org/10.14245/ns.1938398.199

Jensen, M., George, M.J., Russell, M.R., & Odgers, C.L. (2019). Young adolescents' digital technology use and mental health symptoms: Little evidence of longitudinal or daily linkages. *Clinical Psychological Science, 7,* 1416-1433. https://doi.org/10.1177/2167702619859336

Johnson, M.W., & Bickel, W.K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior, 77,* 12-146. https://doi.org/10.1901/jeab.2002.77-129

Johnson, M.W., & Bickel, W.K. (2008). An algorithm for identifying nonsystematic delay-discounting data. *Experimental and Clinical Psychopharmacology, 16,* 264–274. https://doi.org/10.1037/1064-1297.16.3.264

Johnson, M.W., & Bruner, N.R. (2012). The sexual discounting task: HIV risk behavior and the discounting of delayed sexual rewards in cocaine dependence. *Drug and Alcohol Dependence, 123,* 15-21. https://doi.org/j.drugalcdep.2011.09.032

Jones, E.E., & Nisbett, R.E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Lawrence Erlbaum Associates, Inc.

Jönsson, J., & Muhonen, T., (2022). Factors influencing the behavior of bystanders to workplace bullying in healthcare - A qualitative descriptive interview study. *Research in Nursing and Health, 45,* 424-432. https://doi.org/10.1002/nur.22228

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition, 7,* 409–411. https://doi.org/10.1016/0010-0277(79)90024-6

Kaplan, B.A., Reed, D.D., & Jarmolowicz, D.P. (2015). Effects of episodic future thinking on discounting: Personalized age-progressed pictures improve risky long-term health decisions. *Journal of Applied Behavior Analysis, 49,* 148-169. https://doi.org/10.1002/jaba.277

Kaplan, B.A., Reed, D.D., & McKerchar, T.L. (2014). Using a visual analogue scale to assess delay, social, and probability discounting of an environmental loss. *The Psychological Record, 64,* 261-269. https://doi.org/10.1007/s40732-014-0041-z

Kazerooni, F., Taylor, S.H., Bazarova, N.N., & Whitlock, J. (2018). Cyberbullying bystander intervention: The number of offenders and retweeting predict likelihood of helping a cyberbullying victim. *Journal of Computer-Mediated Communication, 00.* 1-17. https://doi.org/10.1093/jcmc/zmy005/4962634

Keane, K.L., & Hammond, S.I. (2022). Interpreting teasing through texting: The role of emoji, initialisms, relationships, and rejection sensitivity in ambiguous SMS. *Canadian Journal of Behavioural Science*. Advance online publication. https://doi.org/10.1037/cbs0000349

Kearney, C.A., & Silverman, W.K. (1995). Family environment of youngsters with school refusal behavior: A synopsis with implications for assessment and treatment. *The American Journal of Family Therapy, 23,* 59-72. https://doi.org/10.1080/01926189508251336

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39,* 1123–1134. https://doi.org/10.1037/0003-066X.39.10.1123

Kirby, M.S., Spencer, T.D, & Spiker, S.T. (2022). Humble behaviorism redux. *Behavior and Social Issues, 31,* 133-158. https://doi.org/10.1007/s42822-022-00092-4

Kirk, B.A., Schutte, N.S., & Hine, D.W. (2008). Development and preliminary validation of an emotional self-efficacy scale. *Personality and Individual Differences, 45,* 432–436. https://doi.org10.1016/j.paid.2008.06.010

Kobus, K. (2003). Peers and adolescent smoking. *Addiction, 98,* 37-55. https://doi.org/10.1046/j.1360-0443.98.s1.4.x

Koehler, C., & Weber, M. (2018). ”Do I really need to help?!” Perceived severity of cyberbullying, victim blaming, and bystanders’ willingness to help the victim. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 12,* Article 4. https://doi.org/10.5817/CP2018-4-4

Komorita, S.S., Parks, C.D., & Hulbert, L.G. (1992). Reciprocity and the induction of cooperation in social dilemmas. *Journal of Personality and Social Psychology, 62,* 607-617. https://doi.org/10.1037/0022-3514.62.4.607

Korndörfer, M., Egloff, B., & Schmukle, S.C. (2015). A large scale test of the effect of social class on prosocial behavior. *PloS one, 10,* e0133193. https://doi.org/10.1371/journal.pone.0133193

Kowalski, R.M., Giumetti, G.W., Schroeder, A.N., & Lattanner, M.R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin, 140,* 1073–1137. https://doi.org/10.1037/a0035618

Kowalski, R.M., Limber, S.P., & McCord, A. (2019). A developmental approach to cyberbullying: Prevalence and protective factors. *Aggression and Violent Behavior, 45,* 20-32. https://doi.org/10.1016/j.avb.2018.02.009

Kraus, M.W., Piff, P.K., & Keltner, D. (2009). Social class, the sense of control, and social explanation. *Journal of Personality and Social Psychology, 97,* 992-1004. https://doi.org/10.1037/0882-7974.13.4.553

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53,* 1017–1031. https://doi.org/10.1037/0003-066X.53.9.1017

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002), Internet Paradox Revisited. *Journal of Social Issues, 58,* 49-74. https://doi.org/10.1111/1540-4560.00248

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D.S., Lin, N, Shablack, H., Jonides, J., & Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PLoS ONE, 8,* e69841. https://doi.org/10.1371/journal.pone.0069841

Kwon, M., Kim, D.-J., Cho, H., & Yang, S. (2013). The Smartphone Addiction Scale: Development and validation of a short version for adolescents. *PLoS One, 8*, e83558. https://doi.org/10.1371/journal.pone.0083558

Lambe, L.J., Hudson, C.C., Craig, W.M., & Pepler, D.J. (2017). Does defending come with a cost? Examining the psychosocial correlates of defending behaviour among bystanders of bullying in a Canadian sample. *Child Abuse & Neglect, 65,* 112-123. https://doi.org/10.1016/j.chiabu.2017.01.012

Lambertz-Berndt, M., & Allen, M. (2017). Confederates. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (Vol 1, pp. 223-256). SAGE Publications, Inc, https://dx.doi.org/10.4135/9781483381411

Lanz, L., Thielmann, I., & Gerpott, F.H. (2021). A social desirability scales desirable? A meta-analytic test of the validity of social desirability scales in the context of prosocial behavior. *Journal of personality, 90,* 203-221. https://doi.org/10.1111/jopy.12662

Lapidot-Leffler, N. (2017). Cybervictims and cyberbystanders in an ethnocultural-religious context: Differences between sectors in Israel. *International Dialogues on Education Journal, 4,* 15-30. https://doi.org/10.53308/ide.v4i3.88

Lassiter, B.J., Bostain, N.S., & Lentz, C. (2021). Best practices for early bystander intervention training on workplace intimate partner violence and workplace bullying.

Journal of Interpersonal Violence, 36, 5813-5837. https://doi.org/10.1177/0886260518807907

Latané, B., & Dabbs, J.M. (1975). Sex, group size and helping in three cities. *Sociometry, 38,* 180-194. https://doi.org/10.2307/2786599

Latané, B., & Darley, J.M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology, 10,* 215-221. https://doi.org/10.1037/h0026570

Latané, B., & Darley J.M. (1970). *The unresponsive bystander why doesn't he help?* Appleton-Century Crofts.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin, 89,* 308-324. https://doi.org/10.1037/0033-2909.89.2.308

Latané, B., & Rodin, J. (1969). A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Experimental Social Psychology, 5,* 189-202. https://doi.org/10.1016/0022-1031(69)90046-8

Lauricella, A.R., Cingel, D.P., Blackwell, C., Wartella, E., & Conway, A. (2014). The mobile generation: Youth and adolescent ownership and use of new media. *Communication Research Reports, 31,* 357-364. https://doi.org/10.1080/08824096.2014.963221

Lawyer, S.R., & Schoepflin, F.J. (2013). Predicting domain-specific outcomes using delay and probability discounting for sexual versus monetary outcomes. *Behavioural Processes, 96,* 71-78. https://doi.org/10.1016/j.beproc.2013.03.001

Lei, H., Mao, W., Cheong, C.M., Wen, Y., Cui, Y., & Cai, Z. (2020). The relationship between self-esteem and cyberbullying: A meta-analysis of children and youth students. *Current Psychology, 39,* 830-842. https://doi.org/10.1007/s12144-019-00407-6

Leung, A.N.M., Wong, N., & Farver, J.M. (2018). You are what you read: The belief systems of cyber-bystanders on social networking sites. *Frontiers in Psychology, 9,* a365. https://doi.org/10.3389/fpsyg.2018.00365

Livingstone, S., Ólafsson, K., Helsper, E.J., Lupiáñez-Villanueva, F., Veltri, G.A., & Folkvord, F. (2017). Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of Communication, 67,* 82–105. https://doi.org/10.1111/jcom.12277

Locey, M.L., & Rachlin, H. (2015). Altruism and anonymity: A behavioral analysis. *Behavioural Processes, 118,* 71-75. https://doi.org/10.1016/j.beproc.2015.06.002

Locey, M.L., Jones, B.A., & Rachlin, H. (2011). Real and hypothetical rewards in self-control and social discounting. *Judgement and Decision Making, 6,* 552-564. https://doi.org/10.1017/S1930297500002515

Loewenstein, G., & Thaler, T.H. (1989). Anomalies: Intertemporal Choice. *Journal of Economic Perspectives, 3,* 181-193. https://doi.org/10.1257/jep.3.4.181

Losby, M.M. (2022). *Emoji Use in Online Communication with Emerging Adults: The Impact on Ambiguity, Affect, and Social Connection* [Doctoral Dissertation], IOWA State University, ProQuest.

Luo, A., & Bussey, K. (2022). Mediating role of moral disengagement in the perpetration of cyberbullying by victims and bystanders. *Journal of Adolescene, 94,* 1142-1149. https:/doi.org/10.1002/jad.12092

Lysenstøen, C., Bøe, T., Hjetland, G.J., & Skogen, J.C. (2021). A review of the relationship between social media use and onine prosocial behavior among adolescents. *Frontiers in Psychology, 12,* 579347. https://doi.org/10.3389/fpsyg.2021.579347

Macaulay, P.J.R., Boulton, M.J. & Betts, L.R. (2019). Comparing early adolescents' positive bystander responses to cyberbullying and traditional bullying: The impact of severity and gender. *Journal of Technology in Behavioral Science, 4,* 253–261. https://doi.org/10.1007/s41347-018-0082-2

Macháčková, H. (2020). Bystander reactions to cyberbullying and cyberaggression: Individual, contextual, and social factors. *Current Opinion in Psychology, 36,* 130-134. https://doi.org/10.1016/j.copsyc.2020.06.003

Macháčková, H., Dedkova, L. & Mezulanikova, K. (2015), Brief report: The bystander effect in cyberbullying incidents. *Journal of Adolescence, 43,* 96-99. https://doi.org/10.1016/j.adolescence.2015.05.010

Macháčková, H., Dědková, L., Sevčíková, A., & Černá, A. (2012). Bystanders' support of cyberbullied schoolmates. *Journal of Community & Applied Social Psychology, 23,* 25-36. https://doi.org/10.1002/casp.2135

Macháčková, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2013). Bystanders' support of cyberbullied schoolmates. *Journal of Community & Applied Social Psychology, 23,* 25-36. https://doi.org/10.1002/casp.2135

Machackova, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2016). Empathic responses by cyberbystanders: The importance of proximity. *Journal of Youth Studies, 19,* 793-804. https://doi.org/10.1080/13676261.2015.1112882

Macháčková, H., & Pfetsch, J. (2016). Bystanders' responses to offline bullying and cyberbullying: The role of empathy and normative beliefs about aggression. *Scandinavian Journal of Psychology, 57,* 169– 176. https://doi.org/10.1111/sjop.12277

Macháčková, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2018). Bystanders' supportive and passive responses to cyberaggression. *Journal of School Violence, 17,* 99-110. https://doi.org/10.1080/15388220.2016.1222499

Marciano, L., Driver, C.C., Schulz, P.J., & Camerini, A.-L. (2022). Dynamics of adolescents' smartphone use and well-being are positive but ephemeral. *Scientific Reports, 12,* 1316. https://doi.org/10.1038/s41598-022-05291-y

Matias, J.N. (2017). *The real name fallacy.* The Coral Project.

Matusiewicz, A.K., Carter, A.E., Landes, R.D., & Yi, R. (2013). Statistical equivalence and test-retest reliability of delay and probability discounting using real and hypothetical rewards. *Behavioural Processes, 100,* 116-122. https://doi.org/10.1016/j.beproc.2013.07.019

Mazur, J.E. (1987). An adjusting procedure for studying delayed reinforcement. In J.E. Mazur, M.L. Commons, J.A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: The effect of delay and of intervening events on reinforcement value* (Vol. 5, pp. 55-73). Erlbaum.

McCarty, S., Teie, S., McCutchen, J. & Geller, E.S. (2016) Actively caring to prevent bullying in an elementary school: Prompting and rewarding prosocial behavior. *Journal of Prevention & Intervention in the Community, 44,* 164-176, https://doi.org/10.1080/10852352.2016.1166809

McDonald, P., Charlesworth, S., & Graham, T. (2016). Action or inaction: Bystander intervention in workplace sexual harassment. *The International Journal of Human Resource Management, 27,* 548-566. https://doi.org/10.1080/09585192.2015.1023331

McKerchar, T.L., Green, L., & Myerson, J. (2010). On the scaling interpretation of exponents in hyperboloid models of delay and probability discounting. *Behavioural Processes, 84,* 440–444. https://doi.org/10.1016/j.beproc.2010.01.003

McKerchar, T.L., Green, L., Myerson, J., Pickford, T.S., Hill, J.C., & Stout, S.C. (2009). A comparison of four models of delay discounting in humans. *Behavioural Processes, 81,* 256-259. https://doi.org/10.1016/j.beproc.2008.12.017

McLoughlin, L.T., Broadhouse, K.M., Clacy, A., Winks, N., Lagopoulos, J., & Hermens, D.F. (2020). The cyberbullying picture series (CyPicS): Developed for use in cyberbullying and cyberbystander research. *Psychological Reports, 124,* 1370-1383. https://doi.org/10.1177/0033294120922488

McMahon, S., & Dick, A. (2011). "Being in a Room with Like-Minded Men": An Exploratory Study of Men's Participation in a Bystander Intervention Program to Prevent Intimate Partner Violence. *The Journal of Men's Studies, 19*, 3-18. https://doi.org/10.3149/jms.1901.3

Mejía-Cruz, D., Green, L., Myerson, J., Morale-Chainé, & Nieto, J. (2016). Delay and probability discounting by drug-dependent cocaine and marijuana users. *Psychopharmacology*, *233*, 2715-2714. https://doi.org/10.1007/s00213-016-4316-8

Mennicke, A., Montanaro, E., Bowling, J., Tirunagari, A., Williams, M., Jules, B. N., Campbell, S., Carlson, H., Farris, S., McClare, V., Kissler, N., McGonagle, A., Pruneda, P., Haley, G., McMahon, S., Correia, C. J., Meehan, E., Benson, J. K., Willard, J., … Sotiroff, A. (2022). A systematic review of validated measures of U.S.-based bystander intervention-related constructs. *Trauma, Violence, & Abuse*, *0*. https://doi.org/10.1177/15248380221137067

Metcalfe, S.N., & Llewellyn, A. (2019). "It's just the thing you do": Physical and digital fields, and the flow of capital for young people's gendered identity negotiation. *Journal of Adolescent Research, 35,* 84-110. https://doi.org/10.1177/0743558419883359

Midgett, A., Doumas, D.M., Johnston, A., Trull, R., & Miller, R. (2017). Rethinking bullying interventions for high school students: A qualitative study. *Journal of Child and Adolescent Counseling, 4,* 146-163. https://doi.org/10.1080/23727810.2017.1381932

Mishna, F., Regehr, C., Lacombe-Duncan, A., Daciuk, J., Fearing, G., & Van Wert, M. (2018). Social media, cyber-aggression and student mental health on a university campus. *Journal of Mental Health, 27,* 222-229. https://doi.org/10.1080/09638237.2018.1437607

Mitchell, S.H., & Wilson, V.B. (2010). The subjective value of delayed and probabilistic outcomes: outcome size matters for gains but not for losses. *Behavioural Processes, 83,* 36-40. https://doi.org/10.1016/j.beproc.2009.09.003

Mittmann, G., Woodcock, K., Dörfler, S., Krammer, I., Pollak, I., & Schrank, B. (2022). "TikTok is my life and Snapchat is my ventricle": A mixed-methods study on the role of online communication tools for friendships in early adolescents. *The Journal of Early Adolescence, 42*, 172-203. https://doi.org/10.1177/02724316211020368

Morgan, D.L. (1996). *Focus groups as qualitative research* (Vol. 16). Sage publications.

Mu, Y., Kitayama, S., Han, S., & Gelfand, M.J. (2015). How culture gets embrained: cultural differences in event-related potentials of social norm violations. *Proceedings of the National Academy of Sciences, 112,* 15348-15353. https://doi.org/10.1073/pnas.1509839112

Mueller, M. (2019, July 23). Challenging the social media moral panic: Preserving free expression under hypertransparency. *Cato Institute Policy Analysis*. https://www.cato.org/policy-analysis/challenging-social-media-moral-panic-preserving-free-expression-under

Myerson, J., & Green, L. (1995). Discounting models of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior, 64,* 263-276. https://doi.org/10.1901/jeab.1995.64-263

Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting, *Journal of the Experimental Analysis of Behavior, 76,* 235-243. https://doi.org/10.1901.jeab.2001.76-235

Necka, E.A., Cacioppo, S., Norman, G.J., & Cacioppo, J.T. (2016). Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. *PLOS One, 11,* e0157732. https://doi.org/10.1371/journal.pone.0157732

Neff, M.B., & Macaskill, A.C. (2021). The effect of "should" and "would" instructions on delay discounting of rewards for self and others. *Behavioral Decision Making, 4,* 568-580. https://doi.org/10.1002/bdm.2230

Neuringer, A. (1991). Humble behaviorism. *The Behavior Analyst, 14,* 1-13. https://doi.org/ https://doi.org/10.1007/BF03392543

Ng, K., Niven, K., & Notelaers, G. (2022). Does bystander behavior make a difference? How passive and active bystanders in the group moderate the effects of bullying exposure. *Journal of Occupational Health Psychology, 27,* 119-135. https://doi.org/10.1037/ocp0000296

Nickerson, A.B., Singleton, D., Schnurr, B., & Collen, M.H. (2014). Perceptions of school climate as a function of bullying involvement. *Journal of Applied School Psychology, 30,* 157-181. https://doi.org/10.1080/15377903.2014.888530

Nikolic, A., Bukurov, B., Kocic, I., Soldatovic, I., Mihajlovic, S., Nesic, D., ... & Grujicic, S.S. (2022). The validity and reliability of the Serbian version of the smartphone addiction scale—short version. *International Journal of Environmental Research and Public Health, 19,* a1245. https://doi.org/10.3390/ijerph19031245

O'Connor, K.M., De Dreu, C.K.W., Schroth, H., Barry, B., Lituchy, T.R., & Bazerman, M.H. (2002). What we want versus what we think we should do: An empirical investigation of intrapersonal conflict. *Journal of Behavioral Decision Making, 15,* 403-418. https://doi.org/10.1002/bdm.426

O'Dea, B., & Campbell, A. (2011). Healthy connections: Online social networks and their potential for peer suport. *Studies in Health Technology and Informatics, 168,* 133-140. https:/doi.org/10.3233/978-1-60750-791-8-133

O'Reilly, M., Dogra, N., Hughes, J., Reilly, P., George, R., & Whiteman, N. (2019). Potential of social media in promoting mental health in adolescents. *Health promotion international, 34,* 981-991. https://doi.org/10.1093/heapro/day056

O'Reilly, M. (2020). Social media and adolescent mental health: the good, the bad and the ugly. *Journal of Mental Health, 29,* 200-206. https://doi.org/10.1080/09638237.2020.1714007

Obermaier, M., Fawzi, N., & Koch, T. (2016). Bystanding or standing by? How the number of bystanders affects the intention to intervene in cyberbullying. *New Media & Society, 18,* 1491-1507. https://doi.org/10.1177/1461444814563519

Odgers, C.L., & Jensen, M.R. (2020). Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. *Journal of Child Psychology and Psychiatry, 61,* 336-348. https://doi.org/10.1111/jcpp.13190

Olenik-Shemesh, D., Heiman, T., & Eden, S. (2015). Bystanders' behavior in cyberbullying episodes: Active and passive patterns in the context of personal-socio-emotional factors. *Journal of Interpersonal Violence, 32,* 23-48. https://doi.org/10.1177/0886260515585531

Olsen, R. (2020). *The effect of episodic future thinking on delay discounting.* [Doctoral thesis, Victoria University of Wellington]. https://doi.org/10.26686/wgtn.17145569.v1

Olsen, R.A., Macaskill, A.C., & Hunt, M.J. (2018). A Measure of Delay Discounting Within the Academic Domain: Delay Discounting and Student Success. *Journal of Behavioral Decision Making, 31,* 522–534. https://doi.org/10.1002/bdm.2074

Olson, E.A., Hooper, C.J., Collins, P., & Luciana, M. (2007). Adolescents' performance on delay and probability discounting tasks: Contributions of age, intelligence, executive functioning, and self-reported externalizing behavior. *Personality and Individual Differences, 43*, 1886-1897. https://doi.org/10.1016/j.paid.2007.06.016

Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education, 12,* 495-510. https:/doi.org/10.1007/BF03172807

Olweus, D., & Limber, S.P. (2018). Some problems with cyberbullying research. *Current Opinion in Psychology, 19, 139*-143. https://doi.org/10.1016/j.copsyc.2017.04.012

Orben, A., Przybylski, A.K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour, 3,* 173–182. https://doi.org/10.1038/s41562-018-0506-1

Palasinski, M. (2012). Online security and cyberbystander relations in mobilzing sex abuse intervention. *Cyberpsychology, Behavior, and Social Networking, 15,* 551-554. https://doi.org/10.1089/cyber.2011.0641

Park, S., Na, E.-Y., & Kim, E.-M. (2014). The relationship between online activities, netiquette and cyberbullying. *Children and Youth Services Review, 42,* 74-81. https://doi.org/10.1016/j.childyouth.2014.04.002

Parris, L., Jungert, T., Thornberg, R., Varjas, K., Meyers, J., Grunewald, S., & Shriberg, D. (2020). Bullying bystander behaviors: The role of coping effectiveness and the moderating effect of gender. *Scandinavian Journal of Psychology, 61,* 38-46. https://doi.org/10.1111/sjop.12564

Paulhus, D.L., & Martin, C.L. (1987). The structure of personality capabilities. *Journal of Personality and Social Psychology, 52,* 354–365. https://doi.org/10.1037/0022-3514.52.2.354

Paull, M., Omari, M., & Standen, P. (2012). When is a bystander not a bystander? A typology of the roles of bystanders in workplace bullying. *Asia Pacific Journal of Human Resources, 50,* 351-366. https://doi.org/10.1111/j.1744-7941.2012.00027.x

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54,* 1643-1662. https://doi.org/10.3758/s13428-021-01694-3

Peña-Fernández, S., Larrondo-Ureta, A., & Morales-i-Gras, J. (2023). Feminism, gender identity and polarization in TikTok and Twitter. *Comunicar*, *21*, 49-60. https://doi.org/10.3916/C75-2023-04

Peters, J., Miedl, S.F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PLoS ONE, 7*, e47225. https://doi.org/10.1371/journal.pone.0047225

Pew Research Center. (2014, February). *The Web at 25.* http://www.pewinternet.org/2014/02/25/the-web-at-25-in-the-u-s

Pew Research Center. (2021, April). *Social Media use in 2021.* https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Phares, E.J., & Wilson, K.G. (1972). Responsibility attribution: Role of outcome severity, situational ambiguity, and internal-external control. *Journal of Personaility, 40*, 392-406. https://doi.org/10.1111/j.1467-6494.1972.tb00069.x

Piff, P.K., Kraus, M.W., Côté, S., Cheng, B.H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology, 99,* 771–784. https://doi.org/10.1037/a0020092

Polanco-Levican, K., & Salvo-Garrido, S. (2021). Bystander roles in cyberbullying: A mini-review of who, how many, and why. *Frontiers in Psychology, 12*, 676787. https://doi.org/10.3389/fpsyg.2021.676787

Poltavski, D.V., & Weatherly, J.N. (2013). Delay and probability discounting of multiple commodities in smokers and never-smokers using multiple-choice tasks. *Behavioural Pharmacology, 24,* 659–667. https://doi.org/10.1097/FBP.0000000000000010

Pouwels, J.L., Keijsers, L., & Odgers, C. (2022). Who benefits most from using social media, the socially rich or the socially poor? *Current Opinion in Psychology, 47,* Article 101351. https://doi.org/10.1016/j.copsyc.2022.101351

Pöyhönen, V., Juvonen, J., & Salmivalli, C. (2012). Standing up for the victim, siding with the bully or standing by? Bystander responses in bullying situations. *Social Development, 21,* 722–741. https://doi.org/10.1111/j.1467-9507.2012.00662.x

Prensky, M. (2001). Digital natives, digital immigrants part 2: Do they really think differently?. *On the horizon, 9,* 1-6. https://doi.org/10.1108/10748120110424843

Price, D., Green, D., & Spears, B., Scrimgeour, M., Barnes, A., Geer, R., & Johnson, B. (2013). A qualitative exploration of cyber-bystanders and moral engagement. *Journal of Psychologists and Counsellors in Schools, 24,* 1-17. https://doi.org/10.1017/jgc.2013.18

Prolific. (2020). Prolific (Version 2.19-2.22) [Online platform]. https://prolific.co

Pyżalski, J., Plichta, P., Szuster, A., & Barlińska, J. (2022). Cyberbullying characteristics and prevention – What can we learn from narratives provided by adolescents and their teachers? *International Journal of Environmental Research and Public Health, 18,* 11589. https://doi.org/10.3390/ijerph191811589

Qin, K., Marien, H., Custers, R., & Aarts, H. (2023). Environmental control of social goals: using Pavlovian-to-instrumental transfer to test cue-based pro-self and pro-social outcome responses. *Royal Society Open Science, 10,* 220660. https://doi.org/10.1098/rsos.220660

Quiroz, S.I., & Mickelson, K.D. (2021). Are online behaviors damaging our in-person connections? Passive versus active social media use on romantic relationships. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15,* a1. https://doi.org/10.5817/CP2021-1-1

Rabb, N., Bowers, J., Glick, D., Wilson, K.H., & Yokum, D. (2022). The influence of social norms varies with “others’ groups: Evidence from COVID-19 vaccination intentions. *Proceedings of the National Academy of Sciences, 119,* e2118770119. https://doi.org/10.1073/pnas.2118770119

Rachlin, H. (1989). *Judgment, decision, and choice: A cognitive/behavioral synthesis*. Times Books.

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior, 85,* 425-435. https://doi.org/10.1901/jeab.2006.85-05

Rachlin, H. (2016). Social cooperation and self-control. *Managerial and Decision Economics, 37,* 249-260. https://doi.org/10.1002/mde.2714

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of Experimental Analysis of Behavior, 55,* 233-244. https://doi.org/jeab.1991.55-233

Rae, J.R., & Lonborg, S.D. (2015). Do motivations for using Facebook moderate the association between Facebook use and psychological well-being? *Frontiers in Psychology, 6,* 138568. https://doi.org/10.3389/fpsyg.2015.00771

Reeves, B., Robinson, T., & Ram, N. (2020). Time for the Human Screenome Project. *Nature, 577,* 314-317. https://doi.org/10.1038/d41586-020-00032-5

Reynolds, B. (2004). Do high rates of cigarette consumption increase delay discounting? A cross-sectional comparison of adolescent smokers and young-adult smokers and nonsmokers. *Behavioural Processes, 67,* 545-549. https://doi.org/10.1016/j.beproc.2004.08.006

Reynolds, B., Richards, J.B., Horn, K., & Karraker, K. (2004). Delay discounting and probability discounting as related to cigarette smoking status in adults. *Behavioural Processes, 65,* 35-42. https://doi.org/10.1016/S0376-6357(03)00109-8

Rice, E.S., Haynes, E., Royce, P., & Thompson, S.C. (2016). Social media and digital technology use among Indigenous young people in Australia: a literature review. *International Journal for Equity in Health, 15,* 18. https://doi.org/10.1186/s12939-016-0366-0

Richards, J.B., Zhang, L., Mitchell, S.H., & De Wit, H. (1999). Delay or probability discounting in a model of impulsive behavior: Effect of alcohol. *Journal of the Experimental Analysis of Behavior, 71,* 121-143. https://doi.org/10.1901/jeab.1999.71-121

Rinderhagen, M., & Sargisson, R.J. (2021). Response efficacy in environmental discounting: Concern and action towards climate change threats. *Sustainability in Environment, 7,* 7-27. http://dx.doi.org/10.22158/se.v7n1p7

Rivers, I., Poteat, V.P., Noret, N., & Ashurst, N. (2009). Observing bullying at school: The mental health implications of witness status. *School Psychology Quarterly, 24*, 211–223. https://doi.org/10.1037/a0018164

Robertson, S.H., & Rasmussen, E.B. (2018). Comparison of potentially real versus hypothetical food outcomes in delay and probability discounting tasks. *Behavioural Processes, 149,* 8-15. https://doi.org/10.1016/j.beproc.2018.01.014

Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education, 5,* 589965. https://doi.org/10.3389/feduc.2020.589965

Roche, B., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I., & O'Hora, D. (2002). Relational frame theory: A new paradigm for the analysis of social behavior. *The Behavior Analyst, 25,* 75-91. https://doi.org/10.1007/BF03392046

Rodríguez-de-Dios, I., van Oosten, J.M.F., & Igartua, J.-J. (2018). A study of the relationship between parental mediation and adolescents' digital skills, online risks and online opportunities. *Computers in Human Behavior, 82*, 186-189. https://doi.org/10.1016/j.chb.2018.01.012

Rosenbaum, G.M., & Hartley, C.A. (2018). Developmental perspectives on risky and impulsive choice. *Philosophical Transactions of the Royal Society B, 374,* a20180133. https://doi.org/10.1098/rstb.2018.0133

Rosenberg M. (1979). *Conceiving the self.* New York: Basic Books.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton: Princeton University Press.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review, 60,* 141–156. https://doi.org/10.2307/2096350

Rosenfeld, M.J., & Thomas, R.J. (2012). Searching for a mate: The rise of the internet as a social intermediary. *American Sociological Review, 77,* 523-547. https://doi.org/10.1177/0003122412448050

Rosenthal, S.R., Buka, S.L., Marshall, B.D.L., Carey, K.B., & Clark, M.A. (2016). Negative experiences on Facebook and depressive symptoms among young adults. *Journal of Adolescent Health, 59,* 510-516. https://doi.org/10.1016/j.jadohealth.2016.06.023

Ross, S.W., & Horner, R.H. (2009). Bully prevention in positive behavior support. *Journal of Applied Behavior Analysis, 42,* 747-759. https://doi.org/10.1901/jaba.2009.42-747

Roussel, S. (2017). Prosocial behaviors. In A. Marciano, & G. B. Ramello (Eds.), *Encyclopedia of law and economics* (pp. 1– 4). Springer. https://doi.org/10.1007/978-1-4614-7883-6_710-1

Rung, J.M., & Madden, G.J. (2018). Experimental reductions of delay discounting and impulsive choice: A systematic review and meta-analysis. *Journal of Experimental Psychology: General, 147,* 1349-1381. https://doi.org/10.1037/xge0000462

Brown, R. (2014, October). Digital media and the internet - The early days of the internet, 1990s. *Te Ara - the Encyclopedia of New Zealand.* http://www.TeAra.govt.nz/en/digital-media-and-the-internet/page-2

Ryff, C.D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069-1081. https://doi.org/10.1037/0022-3514.57.6.1069

Safin, V., Arfer, K.B., & Rachlin, H. (2015). Reciprocation and altruism in social cooperation. *Behavioural Processes, 116,* 12-16. https://doi.org/10.1016/j.beproc.2015.04.009

Sagioglou, C., & Greitemeyer, T. (2014). Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it. *Computers in Human Behavior, 35,* 359-363. https://doi.org/10.1016/j.chb.2014.03.003

Salmivalli, C. (2002). Proactive and reactive aggression among school bullies, victims, and bully-victims. *Aggressive Behavior, 28,* 30-44. https://doi.org/10.1002/ab.90004

Sargisson, R.J., & Schöner, B.V. (2020). Hyperbolic discounting with environment outcomes across time, space, and probability. *The Psychological Record, 70,* 515-527. https://doi.org/10.1007/s40732-019-00368-z

Sargisson, R.J., Ankoné, B.A.H., Dijkhuis, M., Eilers, H., Moser, E., Philipp, A.J., & Syafika, A. (2021). By-and-buy: The effect of social influence on discounting of environmental outcomes. *Sustainability and Climate Change, 14,* 217-226. https://doi.org/10.1089/scc.2021.0003

Sarmiento, A., Herrera-López, M., & Zych, I. (2019). Is cyberbullying a group process? Online and offline bystanders of cyberbullying act as defenders, reinforcers and outsiders. *Computers in Human Behavior, 99,* 328-334. https://doi.org/10.1016/j.chb.2019.05.037

Scharkow, M. (2016). The accuracy of self-reported internet use - a validation study using client log data. *Communication Methods and Measures, 10*, 13-27. https://doi.org/10.1080/19312458.2015.1118446

Schlosnagle, L. (2011). *Following advice because it's been paid for: Age, the sunk-cost fallacy, and loss aversion* [Doctoral Thesis, West Virginia University Graduate]. https://doi.org/10.33915/etd.4783

Schlund, M.W., Carter, H., Cudd, G., Murphy, K., Ahmed, N., Dymond, S., & Tone, E.B. (2021). Human social defeat and approach-avoidance: Escalating social-evaluative threat and threat of aggression increases social avoidance. *Journal of the Experimental Analysis of Behavior, 115*, 157-184. https://doi.org/10.1002/jeab.654

Schønning, V., Hjetland, G.J., Aarø, L.E., & Skogen, J.C. (2020). Social media use and mental health and well-being among adolescents–a scoping review. *Frontiers in psychology, 11,* 1949. https://doi.org/10.3389/fpsyg.2020.01949

Schultze-Krumbholz, A., Hess, M., Pfetsch, J., & Scheithauer, H. (2018). Who is involved in cyberbullying? Latent class analysis of cyberbullying roles and their associations with aggression, self-esteem, and empathy. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 12,* Article 2. https://doi.org/10.5817/CP2018-4-2

Schuster, A.M., Cotten, S.R. & Meshi, D. (2023). Established adults, who self-identify as smartphone and/or social media overusers, struggle to balance smartphone use for personal and work purposes. *Journal of Adult Development, 30,* 78–89. https://doi.org/10.1007/s10804-022-09426-3

Seo, D. G., Park, Y., Kim, M. K., & Park, J. (2016). Mobile phone dependency and its impacts on adolescents' social and academic behaviors. *Computers in human behavior, 63,* 282-292. https://doi.org/10.1016/j.chb.2016.05.026

Shapka, J.D., Onditi, H.Z., Collie, R.J., & Lapidot-Lefler, N. (2018). Cyberbullying and cybervictimization within a cross-cultural context: A study of Canadian and Tanzanian adolescents. *Child Development, 89,* 89-99. https://doi.org/10.1111/cdev.12829

Sheeran, P. (2002). Intention-behaviour relations: A conceptual and empirical review. *European Review of Social Psychology, 12,* 1-36. https://doi.org/10.1080/14792772143000003

Sheeran, P., Harris, P.R., & Epton, T. (2014). Does heightening risk appraisals change people's intentions and behavior? A meta-analyis of experimental studies. *Psychological Bulletin, 140*, 511-543. https://doi.org/10.1037/a0033065

Sidman, M. (1955). On the persistence of avoidance behavior. *Journal of Abnormal and Social Psychology, 50,* 217-220. https://doi.org/10.1037/h0039805

Skinner, B.F. (1953). *Science and human behavior.* Macmillan.

Skinner, B.F. (1971). *Beyond freedom and dignity* (1st ed.). Knopf.

Small, D.A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty, 26*, 5-16. https://doi.org/10.1023/A:1022299422219

Smith, C.L., & Hantula, D.A. (2008). Methodological considerations in the study of delay discounting in intertemporal choice: A comparison of tasks and modes. *Behavior Research Methods 40,* 940–953. https://doi.org/10.3758/BRM.40.4.940

Smith, K.R., Lawyer, S.R., & Swift, J.K. (2018). A meta-analysis of nonsystematic responding in delay and probability reward discounting. *Experimental and Clinical Psychopharmacology, 26,* 94-107. https://doi.org/10.1037/pha0000167

Soldatova, G.U., Rasskazova, E.I., & Chigarkova, S.V. (2020). Digital socialization of adolescents in the Russian Federation: Parental mediation, online risks, and digital competence. *Psychology in Russia: State of the Art, 13,* 191-206. https://doi.org/10.11621/pir.2020.0413

Solomon, L.Z., Solomon, H., & Maiorca, J. (1982). The effects of bystander's anonymity, situational ambiguity, and victim's status on helping. *The Journal of Social Psychology, 117,* 285-294. https://doi.org/10.1080/00224545.1982.9713438

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science, 32,* 1492-1512. https://doi.org/10.1287/mnsc.32.11.1492

Stamos, A., Lange, F., Huang, S.-C., & Dewitte, S. (2020). Having less, giving more? Two preregistered replications of the relationship between social class and prosocial behavior. *Journal of Research in Personality, 84,* e103902. https://doi.org/10.1016/j.jrp.2019.103902

Stark, L. (2010). The science of ethics: Deception, the resilient self, and the APA code of ethics, 1966-1973. *Journal of the History of the Behavioral Sciences, 46,* 337-370. https://doi.org/10.1002/jhbs.20468

Stellar, J.E., Manzo, V.M., Kraus, M.W., & Keltner, D. (2012). Class and compassion: Socioeconomic factors predict responses to suffering. *Emotion, 12,* 449–459. https://doi.org/10.1037/a0026508

Stewart, N., Scheibehenne, B., & Pachur, T. (2018). Psychological parameters have units: A bug fix for stochastic prospect theory and other decision models. Preprint. https://doi.org/10.31234/osf.io/qvgcd

Sticca, F., & Perren, S. (2013). Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. *Journal of Youth and Adolescence, 42,* 739–750. https://doi.org/10.1007/s10964-012-9867-3

Sun, X., Ram, N., Reeves, B., Cho, M.-J., Fitzgerald, A., & Robinson, T. N. (2023). Connectedness and independence of young adults and parents in the digital world: Observing smartphone interactions at multiple timescales using Screenomics. *Journal of Social and Personal Relationships, 40,* 1126–1150. https://doi.org/10.1177/02654075221104268

Sung-min, S., & Byoung-jin, J. (2022). Effects of smartphone overdependence and the quality of friendship on depression among high school students. *Occupational Therapy International, 2022,* 1932326. https://doi.org/10.1155/2022/3932326

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33,* 1-39. https://psycnet.apa.org/doi/10.1146/annurev.ps.33.020182.000245

Tankard, M.E., & Paluck, E.L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review, 10,* 181-211. https://doi.org/10.1111/sipr.12022

Tateno, M., Teo, A.R., Ukai, W., Kanazawa, J., Katsuki, R., Kubo, H., & Kato, T.A. (2019). Internet addiction, smartphone addiction, and hikikomori trait in Japanese young adult: Social isolation and social network. *Frontiers in Psychiatry, 10,* 441708. https://doi.org/10.3389/fpsyt.2019.00455

Thomas, H.J., Connor, J.P., Baguley, C.M., & Scott, J.G. (2016). Two sides to the story: Adolescent and parent views on harmful intention in defining school bullying. *Aggressive Behavior, 43,* 353-363. https://doi.org/10.1002/ab.21694

Thornberg, R., Landgren, L., & Wiman, E. (2018). 'It depends': A qualitative study on how adolescent students explain bystander intervention and non-intervention in bullying situations. *School Psychology International, 39,* 400-415. https://doi.org/10.1177/0143034318779225

Tian, X. (2016). Network domains in social networking sites: Expectations, meanings, and social capital. *Information, Communication & Society, 19*, 188-202. https://doi.org/10.1080/1369118X.2015.1050051

Tian, M., Tao, R., Zheng, Y., Zhang, H., Yang, G., Li, Q., & Liu, X. (2018). Internet gaming disorder in adolescents is linked to delay discounting but not probability discounting. *Computers in Human Behavior, 80,* 59-66. https://doi.org/j.chb.2017.10.018

Tromholt, M., Lundby, M., Andsbjerg, K., & Wiking, M. (2015). The Facebook experiment: does social media affect the quality of our lives. *The Happiness Institute.* http://www. Happinessresearchinstitute.com/publications/4579836749

Twenge, J.M. (2017, September). Have smartphones destroyed a generation? *The Atlantic.* https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/

Twenge, J.M., Martin, G.N., & Campbell, W.K. (2018). Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology. *Emotion, 18*, 765-780. https://doi.org/10.1037/emo0000403

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition, 6,* 13. https://doi.org/10.5334/joc.259

Uram, P., & Skalski, S. (2022). Still logged in? The link between Facebook sddiction, FoMO, self-esteem, life satisfaction and loneliness in social media users. *Psychological Reports, 125*, 218–231. https://doi.org/10.1177/0033294120980970

Utz, S., & Breuer, J. (2016) Informational benefits from social media use for professional purposes: Results from a longitudinal study. *Cyberpsychology Journal of Psychosocial Research in Cyberspace, 10,* 3. https://doi.org/10.5817/CP2016-4-3

Vaillancourt, T., McDougall, P., Hymel, S., Krygsman, A., Miller, J., Stiver, K., & Davis, C. (2008). Bullying: Are researchers and children/youth talking about the same thing? *International Journal of Behavioral Development, 32*, 486-495. https://doi.org/10.1177/0165025408095553

Valkenburg. P.M. (2021). The effects of social media use on adult well-being and ill-being: what we know and what we need to know. Preprint. https://doi.org/10.31234/osf.io/nxq8j.

Valkenburg, P.M., van Driel, I.I., & Beyens, I. (2021a). The Associations of active and passive social media use with well-being: A critical scoping review. Preprint. https://doi.org/10.31234/osf.io/j6xqz

Valkenburg, P.M., Beyens, I., Pouwels, J.L., van Driel, I.I., & Keijsers, L. (2021b). Social media use and adolescents' self-esteem: Heading for a person-specific media effects paradigm. *Journal of Communication, 71,* 56–78. https://doi.org/10.1093/joc/jqaa039

Valle-Jones, J.C., Walsh, H., O'Hara, J., O'Hara, H., Davey, N.B., & Hopkin-Richards, H. (1992). Controlled trial of back support ('lumbotrain') in patients with non-specific low back pain. *Current Medical Research and Opinion, 12*, 604-613. https://doi.org/10.1185/03007999209111527

Van Cleemput, K., Vandebosch, H., & Pabian, S. (2014). Personal characteristics and contextual factors that determine “helping,” “joining in,” and “doing nothing” when witnessing cyberbullying. *Aggressive Behavior, 40*, 383-396. https://doi.org/10.1002/ab.21534

Van Heugten, K. (2011). Theorizing active bystanders as change agents in workplace bullying of social workers. *Families in Society, 92,* 219-224. https://doi.org/10.1606/1044-3894.4090

Vannucci, A., Flannery, K.M., & Ohannessian, C.M. (2017). Social media use and anxiety in emerging adults. *Journal of Affective Disorders, 207*, 163-166. https://doi.org/10.1016/j.jad.2016.08.040

Vannucci, A., Ohannessian, C. M., & Gagnon, S. (2019). Use of multiple social media platforms in relation to psychological functioning in emerging adults. *Emerging Adulthood, 7,* 501-506. https://doi.org/10.1177/2167696818782309

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review, 11,* 274-302. https://doi.org/10.1111/sipr.12033

Vincent, B.T., & Stewart, N. (2020). The case of muddled units in temporal discounting. *Cognition, 198,* 104203. https://doi.org/10.1016/j.cognition.2020.104203

Wakker, P.P. (2010). *Prospect theory: For risk and ambiguity.* Cambridge University Press.

Wang, H., Sigerson, L., & Cheng, C. (2018). Digital nativity and information technology addiction: Age cohort versus individual difference approaches. *Computers in Human Behavior, 90*, 1-9. https://doi.org/10.1016/j.chb.2018.08.031

Wang, J.-L., Jackson, L.A., Gaskin, J., & Wang, H.-Z. (2014). The effects of social networking site (SNS) use on college students' friendship and well-being. *Computers in Human Behavior, 37,* 229-236. https://doi.org/10.1016/j.chb.2014.04.051

Wartella, E., Rideout, V., Montague, H., Beaudoin-Ryan, L., & Lauricella, A. (2016). Teens, health and technology: A national survey. *Media and communication, 4,* 13-23. https://doi.org/10.17645/mac.v4i3.515

Water, E., Cillessen, A.H.N., & Scheres, A. (2014). Distinct age-related differences in temporal discounting and risk taking in adolescents and young adults. *Child Development, 85,* 1881-1897. https://doi.org/10.1111/cdev.12245

Weatherly, J.N., & Derenne, A. (2011). Comparing delay discounting rates when using the fill-in-the-blank and multiple-choice methods. *The Journal of General Psychology, 138*, 300-318. https://doi.org/10.1080/00221309.2011.606442

Weatherly, J.N., Miller, K., & McDonald, T.W. (1999). Social influence as stimulus control. *Behavior and Social Issues, 9,* 25-45. https://doi.org/10.5210/bsi.v9i1.135

White, T. J., Redner, R., Skelly, J. M., & Higgins, S. T. (2015). Examination of a recommended algorithm for eliminating nonsystematic delay discounting response sets. *Drug and alcohol dependence, 154,* 300-303. https://doi.org/10.1016/j.drugalcdep.2015.07.011

Wilcockson, T.D.W., Ellis, D.A., & Shaw, H. (2018). Determining typical smartphone usage: what data do we need? *Cyberpsychology, Behavior, and Social Networking, 21*, 395-398. https://doi.org/10.1089/cyber.2017.0652

Williams, D. (2006). On and off the ‘net: scales for social capital in an online era. *Journal of Computer-Mediated Communication, 11,* 593-628. https://doi.org/10.1111/j.1083-6101.2006.00029.x

Williams, K.D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods, 38,* 174-180. https://doi.org/10.3758/bf03192765

Winskel, H., Kim, T.-H., Kardash, L., & Belic, I. (2019). Smartphone use and study behavior: A Korean and Australian comparison. *Heliyon, 5,* e02158. https://doi.org/10.1016/j.heliyon.2019.e02158

Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs, 59*, 329-349. https://doi.org/10.1080/03637759209376276

Wolter, L.C., & Napper, L.E. (2023). (Mis)perceptions of bystander behaviors and approval: The role of gender-specific perceived norms. *Psychology of Violence*. Advance online publication. https://doi.org/10.1037/vio0000468

Worsteling, A., & Keating, B.W. (2022). Community and bystander interventions for the prevention of suicide: Protocol for a systematic review. *PLoS one, 17,* e0270375. https://doi.org/10.1371/journal.pone.0270375

Wright, M.F. (2013). The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. *Cyberpsychology, Behavior, and Social Networking, 16,* 858-862. https://doi.org/10.1089/cyber.2013.0009

Yoon, J.H., De La Garza, R., Newton, T.F., Suchting, R., Weaver, M.T., Brown, G.S., ... & Haliwa, I. (2017). A comparison of Mazur's k and area under the curve for describing steep discounters. *The Psychological Record, 67,* 355-363. https://doi.org/10.1007/s40732-017-0220-9

Zhao, Y., Chu, X., & Rong, K. (2023). Cyberbullying experience and bystander behavior in cyberbullying incidents: The serial mediating roles of perceived incident severity and empathy. *Computers in Human Behavior, 138,* 107484. https://doi.org/10.1016/j.chb.2022.107484

Zhu, C., Huang, S., Evans, R., & Zhang, W. (2021). Cyberbullying among adolescents and children: A comprehensive review of the global situation, risk factors, and preventative measures. *Frontiers in Public Health, 9,* 634909. https://doi.org/10.3389/fpubh.2021.634909

Ziegler, F.V., & Tunney, R.J. (2012). Decisions for others become less impulsive the further away they are on the family tree. *PLOS ONE, 7,* e49479. https://doi.org/10.1371/journal.pone.0049479

Appendix A: Focus Group Additional Information

Online Registration of Interest Form

General Description of the Study

I am a Ph.D. student studying at the Tauranga branch of the University of Waikato. My research is about how people behave online and what shapes their experiences of online environments. In academic literature, there is this concept of "Digital Natives" which describes people born in 1990 or later. I am interested in what people in that age group think about the internet and online communication, because they're really the experts on this stuff.

Participation in the focus group discussions

Do you remember Bebo? Probably not. That's why the first phase of my research is going to be exploratory focus groups with people like you, so that people like me can research things like cyberbullying in more realistic ways. What I am hoping to learn from the focus groups are things like which social networking sites you use most often, how those sites are being used, and which features of those sites are most useful to you. I am also interested in how different people interpret online comments, so will bring some YouTube® comments to discuss as well. If you have any questions about the study so far, you can include them in this form or email me on ehodge@waikato.ac.nz

About You

Name: ____

Birth Date: __/__/____

Gender Identity: ____________________________________

Best Form of Contact (txt/phone/email): ____

How much time do you spend online (including via phone, Xbox, etc) in the average day?

About ____ hours

How much time do you spend on social networking sites in the average day?

About ____ hours

Anything else you'd like to add? Questions?____

____

Pre-Planned Parts of the Focus Groups

While the majority of the focus group session will be aimed at facilitating a free-flowing discussion led by the participant's experiences, it is important to have a structured introduction and conclusion to address any ethical concerns and relay important information to the participants clearly.

Scripted Guide for Focus Group Introduction

"Thanks for agreeing to participate. My name is Emma-Leigh and I'm a PhD student at the University of Waikato interested in people's behaviour online.

These focus groups are the first practical part of my research project because it's really important that any experiments we design are actually realistic. So, that's where you all come in. There's a whole range of people that use social networking sites on pretty much a daily basis, but "young people" – academics call us digital natives – generally use online communication more and pick up new trends faster than others. Basically, you're the experts here.

And remember that it's totally fine to choose not to answer any questions or to stop participating if you want. You'll also know from the info sheets that we're recording today but that I am the main one who'll hear those back as I transcribe and anonymise them, giving you all code names after that. Does that all sound okay to everyone? Any questions so far?

Cool. Feel free to ask any questions as we go along too.

So I thought we'd start by doing a round introducing ourselves – by a nickname if you like – and saying which social networking sites you use in an average week"

Scripted Guide for Focus Group Conclusion

"Great. Thanks everyone, I think I have plenty to take away and keep working on now.

So my next step is to transcribe and anonymise this discussion, then think through what some of the main themes were. If you'd like to check that I've covered everything accurately, just let me know – my details are on the info sheet – and I can send out copies to you for that.

You've also got a copy of some of the support people and organisations, like Netsafe and Youthline, that are available to you on the info sheet.

Do you have any other questions for me at that moment?

Cool, well thanks again for helping out today"

Questions & Prompts (if required)

The following is a list of questions or prompts that I could use if the discussion falters at any stage:

What do you think is the "best" SNS? What do you think makes it the best?

Okay, so we've got a good list of SNS you all use…can we compare [SNS] X with Y?

Can you think of any features of [SNS] X that you hardly ever use? So for example, I've never used the Up/Down-vote buttons on YouTube.

Do you use one main SNS, or several for different kinds of communication? Why is that do you think?

What's the newest SNS site you've used?

Do you hear the whole "Just don't read the comments" saying very much? What do you think of that?

Are there some SNS you use where people behave worse than on others?

Would you say you mostly communicate with people you know in-person while online? If not, are there any differences in how you talk to people you just know online – can you give an example?

Do you think people talk to one another online the same way they do face-to-face? (Either way, then ask) Can you give an example of that?

Prompts: YouTube® Comments

The following are ten verbatim YouTube® comments (or series of comments) that I collected as part of another research project (see Hodge & Sargisson, 2022) included in the focus groups as a vector for participants sharing their own stories or approaches to harmful online behaviour:

1		"man if you wanna learn to shuffle go to hip hop classes with the girls haahhaha"
2	A	"gangnam style is korean and he's chinese." (Correcting a previous commenter)
	B	"asains, they take things seriously"
	C	"we always have to mention someones race. Please, get a life, if ur gonna just add to the nonsense"
3		"This kid is going to grow up and be one of the best breakdancers. :D"
4		"He's asian so I'm not impressed."
5		"[at Marwa Awil] Shut your stupid face. Why are you so stupis? Cant you see he lied all the way, the fucking deformed chimp" (A clip from a talent show, like America's Got Talent-type show)
6		"The whole dance in general looks off from the beat? It kind of loks like someone slapped this song over the original."
7		"lets see you dance Stupid, shes 9 years old give her a break
8		"She belongs in the movie step up 3"
9		"um I think that u didn't do much dance but did more tricks then dance and u didn't flow with it u stped [sic] and started but u did really well :)"
10		"hey kevin, man you're great for when you were just 11. don't get all bothered by people's negative comments. You just need to know that you are good and the rest just comes. So basically - screw people like Ohm. He most likely cant do even 10% as good as that."

Appendix B: Study 4.1 Verbatim Discounting Scenarios and Questions

Monetary Consequence, High-Severity Scenario (MCH):

Imagine that you have heard through colleagues that an online scam is going around.
The topic comes up in a group conversation and you learn that people are targeted via their public social media profiles, in a very specific way that you all discuss in detail. It is less clear how they scam their targets out of money, though you have heard of some people losing as much as $500.

A few days later you see a social media post that is directed at a friend of yours.

Monetary Consequence, Low-Severity Scenario (MCL):

Imagine that one of your friends has been struggling to promote their on-the-side business creating unique craft projects to order.
Your friend sells their work for a profit usually between $20-$50.
Recently they have been using social media to raise the profile of the business, with a particular focus on displaying reviews from previous customers to establish the quality of their work.

You happen to be on social media when a new review is posted publicly.

Emotional/Social Consequence, High-Severity Scenario (ECH):

Imagine that you are online when you come across a nasty rumour about a close friend of yours.
A recent ex-partner of theirs has publicly posted on social media, accusing your friend of treating them badly and being wholly responsible for the breakdown of the relationship.

They have posted publicly so the accusations are easily viewable to your friends, as well as all their friends and family.

Emotional/Social Consequence, Low-Severity Scenario (ECL):

Imagine that you see a social media post in which someone appears to have an issue with one of your close friends.
Your friend is being accused of being rude for something you know they could not have done.

Knowing your friend, they would be embarrassed to know that they had potentially offended someone, let alone that their own friends and family could come across the accusation.

Example Discounting Trial & Manipulation Check:

You are about 95% sure that the post is another attempted scam.

How likely are you to protect your friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Great, that wraps up this scenario!

Please choose which of the below statements best describes the scenario you just completed, then click "Continue" to go onto the next page.

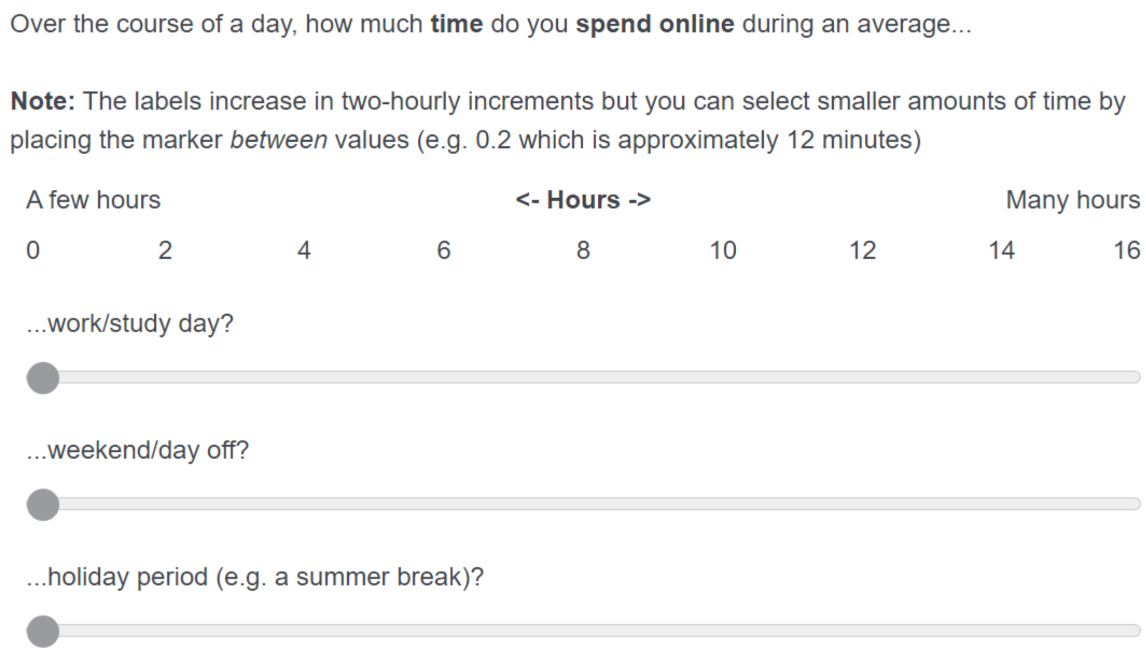
- Someone's ex-partner lashing out
- A potential $500 scam
- Your friend being accused of being rude
- A friend's craft business

Additional Measures:

I included the additional measures of time spent online (TO), time spent on social media (ToSM), self-esteem (RSE), perceived socioeconomic status (PSES), age, gender identity, and study/work status. All addition measures questions and scales are listed in Appendix C.

Appendix C: Additional Measurements

Time Spent Online:



Over the course of a day, how much **time** do you **spend online** during an average...

Note: The labels increase in two-hourly increments but you can select smaller amounts of time by placing the marker *between* values (e.g. 0.2 which is approximately 12 minutes)

A few hours **<- Hours ->** Many hours

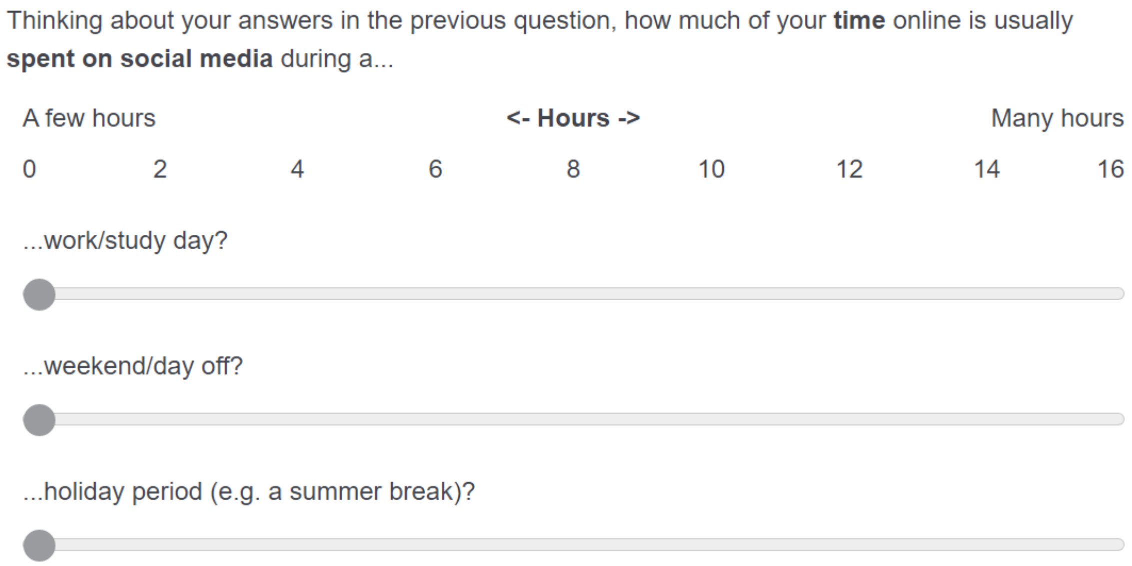
0 2 4 6 8 10 12 14 16

...work/study day?

...weekend/day off?

...holiday period (e.g. a summer break)?

Time Spent on Social Media:



Thinking about your answers in the previous question, how much of your **time** online is usually **spent on social media** during a...

A few hours **<- Hours ->** Many hours

0 2 4 6 8 10 12 14 16

...work/study day?

...weekend/day off?

...holiday period (e.g. a summer break)?

Rosenberg Self-Esteem Scale (RSE):

I included the Rosenberg self-esteem (RSE; Rosenberg, 1965) scale an additional measure in Studies 4.1, 5.1, 5.2, and 6.1. In testing the reliability and validity of the RSE, Fleming and Courtney (1984) found good test-retest reliability ($\alpha = .82$) and internal consistency (α = .88). Moreover, the RSE has been applied successfully with samples crowdsourced using online platforms (Adams et al., 2020), as I have done for many of my discounting studies.

The questions included in the RSE scale all use same response options: strongly agree, agree, disagree, and strongly disagree. A high RSE-scale score indicates higher levels of self-esteem. For standard RSE items, strongly agree = 4; agree = 3; disagree = 2; and, strongly disagree = 1. Several items are reverse coded to identify potential satisficing, and these items were scored: strongly agree = 1; agree = 2; disagree = 3; and strongly disagree = 4. Each individual item (listed below) was randomised and prefaced with the same instruction, “For the following statement, please rate to what extent you agree”.

1	On the whole, I am satisfied with myself	Standard item
2	At times I think I am no good at all	Reverse-coded item
3	I feel that I have a number of good qualities	Standard item
4	I am able to do things as well as most other people	Standard item
5	I feel I do not have much to be proud of	Reverse-coded item
6	I certainly feel useless at times	Reverse-coded item
7	I feel that I am a person of worth	Standard item
8	I wish I could have more respect for myself	Reverse-coded item
9	All in all, I am inclined to think that I am a failure	Reverse-coded item
10	I take a positive attitude toward myself	Standard item

MacArthur Subjective Social Status Scale (PSES):

Think of **a ladder with 10 steps representing where people stand in the country you reside**.

At step 10 are the people who are the best off – those who have the most money, the most education, and the most respected jobs.
At step 1 are the people who are worst off – those who have the least money, least education, and the least respected jobs or no job.

Where would you place yourself on this ladder?

People who are worst off									**People who are best off**
1	2	3	4	5	6	7	8	9	10

Smartphone Addiction Scale - Short Version (SAS-SV)

The smartphone addiction scale (SAS) was developed and initially validated by Kwon et al. (2013; α = .91). Additionally, the short version of the SAS has been validated with several different cultural groupings (e.g., Wang et al., 2018; α = .83), though not yet with Australasian participants.

Higher scores on the SAS indicate that an individual has a greater reliance on their smartphone. Put another way, someone who scores highly on the SAS may demonstrate smartphone addiction. Item responses in the SAS are coded such that 1 = strongly disagree; 2 = disagree; 3 = weakly disagree; 4 = weakly agree; 5 = agree; and, 6 = strongly agree. Each individual item (listed below) was randomised and prefaced by the instruction to "please rate to what extent the statement describes you".

1 Missing planned work due to smartphone use

2 Having a hard time concentrating in class, while doing assignments, or while working due to smartphone use

3 Feeling pain in the wrists or at the back of the neck while using a smartphone

4 Won't be able to stand not having a smartphone

5 Feeling impatient and fretful when I am not holding my smartphone

6 Having my smartphone in my mind even when I am not using it

7 I will never give up using my smartphone even when my daily life is already greatly affected by it

8 Constantly check my smartphone so as not to miss conversations between other people on Twitter, Facebook, or a similar site

9 Using my smartphone longer than I had intended

10 The people around me tell me that I use my smartphone too much

Demographic Questions:

Which of the following best describes **the gender you identify with**?

- ○ Male
- ○ Female
- ○ Transgender
- ○ Intersex
- ○ My gender identity is not listed here
- ○ I would prefer not to answer
- ○ Non-binary

I am currently...

Please select all that apply.

- ☐ Retired
- ☐ A School Student
- ☐ Working Part Time
- ☐ Working Full Time
- ☐ Unemployed
- ☐ Studying Full Time
- ☐ Studying Part Time

In what **year were you born**?

Appendix D: Study 5.1 Verbatim Discounting Scenarios and Questions

Personal Responsibility Conditions, Precursory Statement:

For the next scenario,
please answer the six (6) questions according to how ***you*** would behave.

Personal Responsibility, High-Severity (P.ECH) Scenario:

Imagine that you are online when you come across a nasty rumour about a close friend of yours.

A recent Ex-partner of your friend has accused them of treating the Ex badly and being wholly responsible for the breakdown of the relationship.

The Ex partner has posted publicly and so the accusations are easily viewable to your friend, as well as all their friends and family.

You are about 95% sure that the accusations are untrue and that, if word got back to your friend, that it would be very distressing to them.

How likely are you to protect your friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the P.ECH Scenario:

Great, that wraps up this scenario!

Please choose which of the below statements best describes the scenario you just completed

- ○ Someone's friend being accused of being rude
- ○ Someone's ex-partner lashing out

Who were you asked to imagine yourself as?

- ○ Yourself
- ○ Zee

Personal Responsibility, Low-Severity (P.ECL) Scenario:

Imagine that you see a social media post where it appears that another user has an issue with a close friend of yours.

Your friend is being accused of being rude because they completely ignored the person posting. The poster of the message apparently had said "Hello" to your friend, trying to be friendly, at a recent party.

Knowing your friend, you're is sure that it would have been accidental and think your friend would be embarrassed to know that they had offended someone.

You are 10% sure that the poster is making up a story to intentionally embarrass your friend.

How likely are you to protect your friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the P.ECL Scenario:

Great, that wraps up this scenario!

Please choose which of the below statements best describes the scenario you just completed

- ○ Someone's ex-partner lashing out
- ○ Someone's friend being accused of being rude

Who were you asked to imagine yourself as?

- ○ Yourself
- ○ Zee

Someone Else's Responsibility Conditions, Precursory Statement:

For the next scenario,
please answer the six (6) questions according to how **a generic *other person*** would behave, e.g. how **an acquaintance** of yours might respond.
For the purpose of this section, imagine that this acquaintance is **named "Zee"**.

Someone Else's Responsibility, High-Severity (O.ECH) Scenario:

Imagine that Zee is online when they come across a nasty rumour about a close friend of theirs.

A recent Ex-partner of Zee's friend has accused them of treating the Ex badly and being wholly responsible for the breakdown of the relationship.

The Ex partner has posted publicly and so the accusations are easily viewable to Zee's friend, as well as all their friends and family.

Zee is about 95% sure that the accusations are untrue and that, if word got back to Zee's friend, that it would be very distressing to them.

How likely is Zee to protect their friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the O.ECH Scenario:

Great, that wraps up this scenario!

Please choose which of the below statements best describes the scenario you just completed

- ○ Someone's ex-partner lashing out
- ○ Someone's friend being accused of being rude

Who were you asked to imagine yourself as?

- ○ Yourself
- ○ Zee

Someone Else's Responsibility, Low-Severity (O.ECL) Scenario:

Imagine that Zee sees a social media post where it appears that another user has an issue with a close friend of theirs.

Zee's friend is being accused of being rude because they completely ignored the person posting. The poster of the message apparently had said "Hello" to Zee's friend, trying to be friendly, at a recent party.

Knowing thier friend, Zee is sure that it would have been accidental and thinks their friend would be embarrassed to know that they had offended someone.

Zee is about 95% sure that the poster is making up a story to intentionally embarass their friend.

How likely is Zee to protect their friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the O.ECL Scenario:

Great, that wraps up this scenario!

Please choose which of the below statements best describes the scenario you just completed

- Someone's friend being accused of being rude
- Someone's ex-partner lashing out

Who were you asked to imagine yourself as?

- Zee
- Yourself

Additional Measures:

I used the same set of additional measures as in Study 4.1 with the addition of the Smartphone Addiction Scale, all of which are presented in Appendix C.

Appendix E: Study 5.2 Verbatim Discounting Scenarios and Questions

Personal Responsibility Conditions, Precursory Statement:

For the next scenario,
please answer the six (6) questions according to how ***you*** would behave.

Personal Responsibility, Large-Audience (PL) Scenario:

Imagine that you are online when you come across a nasty rumour about a close friend of yours.

A recent ex-partner of theirs has posted on social media, accusing your friend of being wholly responsible for the breakdown of the relationship.

They have posted publicly to a space where your friend is all-but guaranteed to see the post, along with upwards of 50 other onlookers.

You are about 5% sure that the accusations are untrue and that your friend will be very distressed by it.

How likely are you to protect your friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the PL Scenario:

Great, that wraps up this scenario! Just a few quick questions about the scenario you just read...

Who were you asked to imagine yourself as?

- ○ Zee
- ○ Yourself

Approximately how many other people did you imagine witnessing the scenario?

- ○ 50 or more people
- ○ One or two people

Personal Responsibility, Samll-Audience (PS) Scenario:

Imagine that you are online when you come across a nasty rumour about a close friend of yours.

A recent ex-partner of theirs has posted on social media, accusing your friend of being wholly responsible for the breakdown of the relationship.

They have posted publicly to a space where your friend is all-but guaranteed to see the post but it is unlikely that others will, perhaps just you and one other onlooker.

You are about 5% sure that the accusations are untrue and that your friend will be very distressed by it.

How likely are you to protect your friend by replying to call-out the person who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the PS Scenario:

Great, that wraps up this scenario! Just a few quick questions about the scenario you just read...

Who were you asked to imagine yourself as?

- ○ Yourself
- ○ Zee

Approximately how many other people did you imagine witnessing the scenario?

- ○ One or two people
- ○ 50 or more people

Someone Else's Responsibility Conditions, Precursory Statement:

For the next scenario,
please answer the six (6) questions according to how **a generic *other person*** would behave, e.g. how **an acquaintance** of yours might respond.
For the purpose of this section, imagine that this acquaintance is **named "Zee"**.

Someone Else's Responsibility, Large-Audience (OL) Scenario:

Imagine that Zee is online when they come across a nasty rumour about a close friend of theirs.

A recent ex-partner of Zee's friend has posted on social media, accusing Zee's friend of being wholly responsible for the breakdown of the relationship.

The ex-partner has posted publicly to a space where Zee's friend is all-but guaranteed to see the post, along with upwards of 50 other onlookers.

Zee is about 5% sure that the accusations are untrue and that their friend will be very distressed by it.

How likely is Zee to protect their friend, just as publicly, by replying to call-out the ex-partner who posted?

(0%) Unlikely to reply Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the OL Scenario:

Great, that wraps up this scenario! Just a few quick questions about the scenario you just read...

Who were you asked to imagine yourself as?

- ○ Zee
- ○ Yourself

Approximately how many other people did you imagine witnessing the scenario?

- ○ One or two people
- ○ 50 or more people

Someone Else's Responsibility, Small-Audience (OS) Scenario:

Imagine that Zee is online when they come across a nasty rumour about a close friend of theirs.

A recent ex-partner of Zee's friend has posted on social media, accusing Zee's friend of being wholly responsible for the breakdown of the relationship.

The ex-partner has posted publicly to a space where Zee's friend is all-but guaranteed to see the post but it is unlikely that others will, perhaps just Zee and one other onlooker.

Zee is about 5% sure that the accusations are untrue and that their friend will be very distressed by it.

How likely is Zee to protect their friend, just as publicly, by replying to call-out the ex-partner who posted?

(0%) Unlikely to reply — Will definitely reply (100%)

0 20 40 60 80 100

Manipulation Check of the OS Scenario:

Great, that wraps up this scenario! Just a few quick questions about the scenario you just read...

Who were you asked to imagine yourself as?

- ○ Yourself
- ○ Zee

Approximately how many other people did you imagine witnessing the scenario?

- ○ One or two people
- ○ 50 or more people

Additional Measures:

I included the TO, ToSM, PSES, RSE, SAS, age, gender identity, and study/work status which are presented verbatim in Appendix C.

Severity-Ranking Activity:

Please rank the following scenarios involving Person X and/or Person Y by dragging and dropping each statement so that those you consider to be **least severe appear at the top** of the list, with **low numbers** (i.e. the *least* severe statement should end up as #1 at the very top).

- X and Y have never gotten along well, then X unfairly posts online to tell all their mutual friends to start avoiding Y
- Y finds out that they have been removed from a Group Chat without warning or explanation
- X posted online to accuse Y of being rude to them at a recent party, even though Y was not present
- On several different online posts of theirs, Y receives off-topic mean comments from X
- Every time X posts a picture online, they mention Y in a mean hashtag
- Y knows that they are not included in the Group Chat that all their peers are part of
- X and Y have recently broken up and Y lashes out by posting intimate content of X online to hurt them
- An ex-partner is spreading untrue and hurtful information about Y online
- Y is spreading untrue and hurtful information about X's recent break-up
- Y has sent explicit content to X several times, even though it is clear X finds it distressing

Appendix F: Study 6.1 Verbatim Discounting Scenarios and Questions

Personal Responsibility Conditions, Precursory Statement:

Please answer the next seven (7) questions according to **how <u>you</u> personally would behave.**

Imagine that you witness the following scenario online…

Someone Else's Responsibility Conditions, Precursory Statement:

Please answer the next seven (7) questions according to **how you think <u>someone else</u> would behave**, e.g., how an acquaintance of yours might behave.
For the purpose of this section, **imagine that this acquaintance is named "Zee".**

From Zee's perspective, imagine witnessing the following scenario online…

Scenario Statements:

Table 19 *Scenario Statements Used in Study 6.1*

ID #	Scenario Statement
1	Y never receives any response from X when they try talking to them during online multiplayer games.
2	Y knows that they are not included in a Group Chat that X created and added all of Y's peers to.
3	X starts a nasty argument with Y in a comments section when a private message would have been less confrontational.
4	X posted online to accuse Y of being rude to them at a recent party, even though Y was not present.
5	An ex-partner (X) is spreading untrue and hurtful information about Y online.
6	X has sent explicit content to Y several times, even though it is clear Y finds it distressing.
7	X has filmed Y without their knowledge, then posted the video online to make fun of them.

Note. The scenario statements are listed from least- (Statement 1) to most severe (Statement 7) according to the ratings given by participants of Study 5.2.

Manipulation Checks:

In the previous seven (7) questions, whose perspective were you being asked to take?

- ◯ Another person's (Zee's) perspective
- ◯ Your own personal perspective

From those below, please select the statement that did *not* appear within the previous seven (7) questions

- ◯ X has sent explicit content to Y several times, even though it is clear Y finds it distressing
- ◯ Y never receives any response from X when they try talking to them during online multiplayer games
- ◯ Y's online friend request to X is ignored
- ◯ An ex-partner (X) is spreading untrue and hurtful information about Y online

Additional Measures:

I used the same set of questions as earlier studies for TO, ToSM, PSES, RSE, SAS, age, gender identity, and study/work status. See Appendix C for further detail.

Adapted (from Study 5.2) Severity Rating Activity:

Please rank the following scenarios by **dragging and dropping** each statement so that those you consider to be **least severe appear at the bottom** of the list and those you consider to be **most severe are at the top**.

Most Severe

- Y never receives any response from X when they try talking to them during online multiplayer games
- An ex-partner (X) is spreading untrue and hurtful information about Y online
- X has filmed Y without their knowledge, then posted the video online to make fun of them
- X posted online to accuse Y of being rude to them at a recent party, even though Y was not present
- Y knows that they are not included in a Group Chat that X created and added all Y's peers to
- X has sent explicit content to Y several times, even though it is clear Y finds it distressing
- X starts a nasty argument with Y in a comments section, when a private message would have been less confrontational

---

Least Severe