Defining Digital Self-Harm

Jessica Pater and Elizabeth Mynatt
School of Interactive Computing, Georgia Tech
85 5th Street NW, Atlanta GA 30332
{pater; mynatt}@gatech.edu

ABSTRACT
Self-harm is the infliction of pain or injury onto oneself. Though historically these behaviors were relegated to the fringes of communities, information technology now enables new ways to foster and encourage these dangerous activities. This paper defines the concept of digital self-harm as the online communication and activity that leads to, supports, or exacerbates, non-suicidal yet intentional harm or impairment of an individual’s physical wellbeing. We outline a research agenda for the CSCW community to understand the correlation and possible causation of offline self-harm behaviors due to online activities, and to design and assess technologies focused on prevention, mitigation and treatment. CAUTION: This paper includes media that could potentially be triggering to those dealing with an eating disorder or with other self-harm related illnesses. Please use caution when reading or disseminating this paper.

Author Keywords
Digital self-harm; self-harm; self-injury; social computing; pro-ana; pro-mia; thinpiration; cutting; computer-mediated communication; social media; social computing; mental health; behavioral health; health

ACM Classification Keywords
K.4.0 Computers and Society - General

INTRODUCTION
Self-harm, the infliction of pain or harm onto oneself [44], is a mental illness that is growing in prevalence. Examples of self-harm behaviors include restrictive and binge eating, cutting, hitting, scratching, burning, bleaching of an individual’s body, and other forms of self-mutilation [32]. Prior to the Internet, individuals seeking to self-harm or seeking information related to tips and tricks, or best practices, associated with these destructive behaviors were often isolated, relegated to the fringes of society [2]. However, with greater connectivity and access provided by information and communication technologies (ICTs), these groups of like-minded individuals have created online spaces to foster and encourage self-harm behaviors and activities. This online interaction is not a new phenomenon - Usenet and bulletin boards gave power to individuals seeking to find like-minded support via the connections it established. Soon after, chat rooms and instant messaging became a way of connecting individuals together to support one another’s eating disorders [20], body dysmorphia [23], and self-mutilation [66].

Now, in many countries, the Internet is no longer a destination that one engages solely from behind the screen of a desktop computer – the pervasiveness of mobile computing means that a large portion of the population is continually connected [35]. Concerning examples of how continuous sustained access to people and content has changed the presentation and impact of negative behaviors include cyberbullying [4] and cyberstalking [56]. CSCW researchers have shown many ways that ICTs support personal expression with a strong focus on positive interactions. Indeed, most models of information practice or behavior skew towards online information-seeking as positive or neutral in nature [20]. Health-focused CSCW and HCI research in particular tends to favor prosocial studies focused on positive or neutral health promotion or outcomes [13,29,31]. We posit that this unintentional research bias is problematic. As we continue to design online social platforms and their capabilities, it is imperative that we understand these biases and the potential impact they have on platform design and policy.

In this paper, we synthesize previous literature from multiple fields to explore the dynamic relationship between online, mobile, and pervasive computing practices and the growing genre of self-harm behaviors that are mediated through the use of online platforms. We define digital self-harm as the online communication and activity that leads to, supports, or exacerbates non-suicidal yet intentional harm or impairment of an individual’s physical wellbeing. Therefore, digital self-harm occurs when a person’s behaviors are negatively influenced through his or her online activities in such a manner that these online activities lead to the infliction of non-suicidal direct or indirect physical harm to oneself. Digital self-harm includes both the consumption and production of online content, and is directly tied to the clinical constructs that frame our definition.
In this paper, we explore how current CSCW research can inform our understanding of digital self-harm and avenues for future research. We begin by reviewing the clinical definitions of self-harm and its characteristics. We use this foundation to define and contextualize the boundaries of digital self-harm - what it does and does not encompass. We show how the current state of CSCW research informs this definition and identify open questions the community is poised to address. To explore these questions, we review a focused set of theoretical and methodological frameworks. We then conclude with a discussion of the importance of this research domain for the CSCW field through the articulation of remaining open questions as well as implications for designing technical and clinical interventions.

In this paper we make the following contributions to the CSCW community:

1. Provide an operational definition for digital self-harm, derived from both the social computing domain and clinical contexts that include behavioral psychology.
2. Highlight theoretical approaches salient for researching digital self-harm behaviors and activities and provide examples of how the approaches can be operationalized for research in the CSCW domain.
3. Organize a research agenda focused on a deeper understanding of the characterization of digital self-harm activities and the possible correlation and causation between digital self-harm activities and offline behaviors.

Ultimately, we believe that the CSCW community has an obligation to pay closer attention to digital self-harm activity. The social computing systems created and studied by our community, in part, enable this subset of unhealthy behaviors. As designers of technology, we should also be aware of how our tools support destructive behaviors to permeate through society and consider steps we can take to address this growing issue.

**CAUTION:** This paper includes media that could potentially be a trigger to those dealing with an eating disorder or with other self-harm related illnesses. Please use caution when reading, printing, or disseminating this paper.

**WHAT IS SELF-HARM?**

Self-harm is a term commonly used to describe a set of behaviors and activities associated with individuals who cause physical pain or injury to oneself [69]. Traditionally, definitions of self-harm focus on explicit direct non-suicidal actions that impact an individual’s health, starting with cutting and wrist slashing [24] to the inclusion of other direct injuring of body tissue [33], and also extending to indirect injuring like that seen with eating disorders [68]. Reasons of why someone would self-harm vary. Some posit that these acts are an extreme way to cope with distressing or painful feelings [69]. Others focus on this disorder being more associated with body dysmorphic delusions [19].

the breadth and depth of self-harm activities, it becomes possible to investigate how technology impacts these behaviors. Self-harm is a clinical behavioral disorder defined in terms of direct and indirect physical injury. Research shows that these types of behaviors (e.g. restrictive and binge eating, cutting, hitting, scratching, burning, and bleaching of an individual’s body) are not just relegated to the individual—they have become an issue of public health or a new “social pathology” [28].

Our research on this topic began with our exposure to the growing popularity of pro-eating disorder resources online and their uptake by adolescent girls [49]. This behavior is particularly problematic in that self-harm behaviors usually manifest during adolescence [3,28]. A 2014 World Health Organization Study survey found that 20% of 15 year-olds surveyed reported having self-harmed within the last 12 months [15]. Eating disorders have the highest mortality rate of all mental illnesses [57], and the mortality rate associated with Anorexia is 12 times higher than the mortality rate associated with all other causes of death for females 15-24 years old [60]. As we discuss throughout this paper, technological affordances have helped exacerbate this mental illness.

It is important to understand that, as with any other mental or behavioral health issues, the clinical presentations of self-harm can vary much more than traditional physical disorders due to the potential co-occurrence of self-harm with other mental health issues such as depression, anxiety, and personality disorders. For our purposes, we advocate focusing on the expression of self-harm behaviors online, with an understanding that this online content may also include indications of related mental health disorders.

**Direct and Indirect Self-harm**

Self-harm behavior can be classified as either direct or indirect self-harm. Below we discuss these distinctions, the scope of behaviors within these groups, and how they form a clinical definition of self-harm behavior. Understanding these distinctions allows the CSCW community to operationalize these behaviors within a socio-technical context – a critical component in conceptualizing and defining digital self-harm.

**Direct self-harm**

Direct self-harm is defined as the deliberate, direct destruction or alteration of body tissue without conscious suicidal intent, but resulting in injury severe enough for tissue damage to occur [17]. Direct self-harm manifests itself in many forms including the cutting, burning, scratching, and hitting of oneself [40]. Other forms of direct self-harm include skin bleaching [12], hair-pulling [30], and the ingestion of hazardous chemicals [25].

Defining direct self-harm is linguistically challenging: it is often used interchangeably with the terms self-injury [32], self-mutilation [18], and self-wounding [61]. The unifying thread across these terms is the distinction that an
individual is non-suicidal and inflicting direct injury to their bodily tissue. In the most recent edition of the Diagnostic and Statistical Manuel of Mental Disorders (DSM-V), non-suicidal self-injury (NSSI) is now recognized as a distinct health condition.

**Indirect self-harm**

Indirect self-harm refers to non-suicidal behavior that results in an individual physically mistreating or abusing him or herself, yet not causing immediate direct bodily damage [59]. Examples of indirect self-harm activities include eating disorder behaviors, involvement in physically abusive relationships, prolonged substance abuse, and risky or reckless behavior [59]. The eating disorder (ED) community tends to be the most illustrative indirect self-harm community and, as such, it is where a majority of the indirect self-harm literature is focused [2,7,20,55].

Indirect self-harm does not have the same presentation as direct self-harm - it typically manifests itself as a form of expression, a way to ground oneself, and a way to feel and regain control and to punish [46]. It can also often be used as a tool or mechanism to help cope with emotional pain, intense anger and frustration [59], and can even be used by some to decrease tension [25].

**Self-Harm vs. Suicide**

A common misperception is that people who self-harm are suicidal [51]. Suicide is defined as the act of intentionally ending one’s own life [45]. Disentangling suicidal behavior from self-harm is complex, as it can often be the ultimate resolution of self-harm behaviors. As Nock et al. point out, clinicians do distinguish between suicidal behavior and non-suicidal self-harm [45] Suicidal behavior and self-harm behavior require distinct treatment and prevention strategies. While both are often treated with cognitive behavior therapy, suicidal behavior is often also treated pharmacologically and with targeted therapies (e.g. CBT-SP, TASA) to address its acute nature [64]. In contrast, treatment for eating disorders generally includes individual and family/group therapy with a focus on the interpersonal and cultural forces that contribute to, or maintain, self-harm activities (e.g. eating disorders).

With an established clinical definition of self-harm, including understanding its forms, parameters, and the depth and magnitude of self-harm as a public health issue, it is now possible to explore how the introduction of communication technologies influences and impacts the formulation of self-harm behaviors within a digital context.

**DEFINING DIGITAL SELF-HARM**

We define digital self-harm as the online communication and activity that leads to, supports, or exacerbates, non-suicidal yet intentional harm or impairment of an individual’s physical wellbeing. The computing field is limited in its examples of scholarship focused in the domain of self-harm presentations online and, as such, lacks a coherent framework of what encompasses digital self-harm. This paper aims to fill that void.

By using the clinical definitions of self-harm in the previous section, the digital variant of self-harm can be extended to capture nuances specific to the online domain. Access to content, people and institutions that support or reinforce destructive, non-suicidal ideas or behaviors potentially drive the degree of impact digital self-harm has on an individual.

By removing barriers of time, distance, and transience, the landscape of the Internet – be it a community, access to information, or the ability to freely express one’s diseased self – creates a fertile environment for the potential deepening and strengthening of these behavioral issues within an affected individual through the sharing and reinforcement of destructive behaviors. This continued reinforcement can potentially lead to a quicker and deeper presentation of the underlying symptoms of these diseases.

As we discuss throughout this paper, there is notable variation in online venues and media that promote self-harm behavior. Computing researchers have begun to study these issues in earnest. While several studies have begun to characterize digital self-harm activities [6,11,20,48,55], none to date provide an understanding of how different online affordances impact various manifestations of self-harm disorders.

**Examples of Digital Self-Harm**

Below we describe two examples of digital self-harm activities: eating disorder communities and cutting communities. These illustrative examples highlight how digital self-harm behavior manifests in popular online communities.

**Eating Disorder (ED) Communities**

ED-focused online communities are collections of websites, blogs, and social media feeds known as pro-ana (or pro-Anorexia) and pro-mia (or pro-Bulimia) [55]. ED communities promote their activities as healthy or “normal” and frame their activities as a legitimate lifestyle option [20]. These communities usually contain supportive and persuasive language, portals to share information and best practices, discussion forums, and media sharing capabilities. Current research in this space focuses on community support [53], community activities [21,43,48], ethical issues [55] and information sharing practices [11,20].

ED communities are interactive in nature. They allow community members to communicate with each other and

---

1. [http://dsm.psychiatryonline.org](http://dsm.psychiatryonline.org)
contribute to the culture of the community by sharing media, resources, and personal experiences [6,20]. Participants in ED communities often use strong and direct language to ensure the makeup of the community share similar beliefs and traits. These communities offer a place to learn from community members away from the scrutiny of family, friends, and the medical system [43]. Below is an example of a post that reaffirms pro-ana norms and is an example of digital self-harm:

IF YOU WANT TO LOSE WEIGHT, GO ON A DIET FATTY. ONE IS EITHER ANA/MIA OR NOT. IT IS A GIFT AND YOU CANNOT DECIDE TO HAVE AN EATING DISORDER. SO IF YOU ARE LOOKING FOR A WAY TO LOSE WEIGHT, S-S-S-SORRY JUNIOR!! MOVE ON, TRY JENNY CRAIG [6].

In addition, several communities provide tools that allow for members to measure food intake, body mass, and activity levels as a way to encourage members to support each other’s restrictive behaviors [6].

Media sharing is an important feature in many of these online communities [48]. The ability to reinforce shared norms within the community is readily done through the sharing of thinspiration, or thinspo for short. This blend of thin and inspiration involves the sharing of photos, videos, and prose that are intended to inspire eating disorders [36]. Image 1 shows a compilation of traditional forms of thinspiration one might encounter in these online communities [48].

Because of the influential and impactful nature of this type of display, these types of media serve as powerful motivators. Additionally, persistence of these exemplars online plus ease of access ensure that whenever someone needs to reconnect to the destructive and unhealthy message it reinforces, it is readily available. Research has shown ED websites may have contagion-like effects on individuals experimenting or seeking information about eating disorders before exhibiting signs and behaviors of commitment to them [6]. With respect to adolescents and eating disorders, one concern is that a relatively healthy, or at least stereotypical, desire to lose weight may lead to unintended exposure to this type of content.

From a U.S. policy standpoint, these communities often highlight a tension between first amendment rights and censorship of content. Webhosts have a historical precedent for taking down sites that promote Pro-ED discourse, further driving these communities and their members underground [55] and further isolating them from the general Internet community [2]. Attempts to systematically “rid” the Internet of this type of community have proven to be futile, as community members have found ways to migrate ISPs [55] and change the shared vernacular used for searching and sharing of information [11], all without diminishing the size and scope of their communities.

**Image 1. Examples of Thinspiration or Thinspo**

**Image 2. Examples of cutting inspiration**

**Cutting Communities**

Online communities that support the act of cutting are also representative of destructive digital self-harm activities. Similar to pro-ana and pro-mia communities, pro-cutting communities offer a platform for the sharing of best practices and tools of the trade. Like those that participate in ED-community activities, individuals who identify with cutting communities often use specific vernacular to identify or attach themselves to the group within the larger masses. These terms include cat(scratch); cut(s)(ing)(ter); razor(s); scar(red); and selfharm(m)(mm) [48].

The intent of the individuals that use these terms can be classified as either normalizing or pathologizing. Normalizing uses of these terms seek to legitimize these activities as a way of dealing with mental health issues whereas pathologizing uses are focused more on an individual’s body being repulsive and the behavior is seen as morally reprehensible. This discourse is carried out through the more common text, images, and video, but also through original artwork, music and other expressive communication mediums. Image 2 is an example of this discourse presented graphically through image.

These examples of eating disorder and cutting communities are illustrative and help articulate the scope of digital self-harm activities. They provide a look at how both indirect (eating disorder) and direct (cutting) self-injury behaviors are presented in online spaces. While there are several similarities between these two examples, there are also distinct differences. One distinct difference is that direct
harm content tends to focus more on action-oriented activities. These distinctions highlight the need to investigate other forms of self-harm activities within these digital spaces.

What is not included
It is important to note that there are other forms of self-harm that are not included in the clinical definition informing our agenda. Self-injury associated with people with developmental disabilities, such as children on the autism spectrum, is an example of self-injury behaviors that is not included in the clinical classification informing our agenda. The most common forms of these behaviors include head banging, excessive self-rubbing and scratching, hand biting, and picking of the skin and/or face [16]. This type of self-injury is not included because they are often involuntary actions that are neurologically triggered by an underlying disability.

Additionally, activities that focus on using online content to incite an individual to inflict pain on others are not included in this definition. The Internet has been used as a vehicle to do just that – encourage individuals to act or behave in a manner that negatively impacts others both online and offline – in the contexts of behaviors such as terrorism, pedophilia, and racism. Thus, if the underlying cause of these behaviors is rooted within a recognized medical neurological condition, or the person targeted for injury is someone other than the perpetrating individual, that behavior is not considered an example of digital self-harm.

Cyberbullying is notable in that the pervasive access to content, coupled with anonymity, significantly changed the experience of bullying for both perpetrators and victims [27]. Based on our framing, cyberbullying behaviors are not classified as digital self-harm. However, looking at the genesis of cyberbullying, its presentations, and how society has begun to address these issues could be useful to our field as we begin to further understand digital self-harm behaviors and issues.

We have defined the concept of digital self-harm, grounding this concept in both the literature and through illustrative examples. These illustrative examples of digital self-harm practices highlight how the infusion of technology into these behaviors have increased their visibility and potentially allowed people to quickly adopt their practices. While we can point to examples of these activities, left unanswered are questions such as whether digital self-harm activities lead to an increased prevalence of self-harm practices overall and the degree to which different forms of online activities differ in their offline influence.

THEORETICAL LENSES
In establishing digital self-harm as a topic of concern, it is important to understand the applicable theories from within the CSCW field and in related fields such as health and communication. In doing so, we are able to harness explanatory power from multiple domains and employ different theoretical underpinnings that drive a richer understanding of these destructive behaviors, as well as identify methods that the community might use to investigate intervention strategies. Digital self-harm is a multifaceted construct. Furthering our understanding of the interplay between technologies, behavioral health, information seeking, and socialization within communities will require researchers to take an integrative theoretical approach.

Theoretical Framing
While there are many theoretical frameworks that could be used to operationalize the idea of digital self-harm, in this paper we draw attention to the potential of Social Cognitive Theory and Cultivation Theory as a way to understand behavior within the context of life on the screen. We selected these theories based on their explanatory power for understanding the connections between peer influence, psychological drives, and technology use. We also want to call attention to the potential of drawing from adjacent fields including cognitive psychology and social communications theory.

Social Cognitive Theory
Social Cognitive Theory (SCT), rooted in cognitive psychology, explains individual behavior change within the context of society-at-large [5]. The theory posits that when an individual observes a specific behavior and its consequences, they remember and use these recalled memories to inform future actions [5]. Therefore, SCT is a tool that can be used to understand different social, psychological, and environmental influences on behavior. SCT also explains how behavior change can be deeply influenced by a person’s belief in his or her own ability to make a change (i.e. self-efficacy) [8]. The power of employing the SCT lens to study digital self-harm lies in its ability to analyze the impacts of accessing digital self-harm content found online. Examples of this research include understanding the influence of thinspiration and interaction online related to engaging in self-harm behaviors.

SCT-influenced therapies or interventions for digital self-harm would therefore look at technology’s influence on the individual. Understanding how an individual uses online platforms to access information, people, or influences related to his or her disease could give clinicians a more holistic understanding of factors influencing different presentations of self-harm. With this expanded level of information, interventions as well as in- and out-patient treatments could be more targeted, and therefore potentially lead to improved health outcomes [1]. SCT could also inform tools to understand how individuals transition from a state of not fully engaging in digital self-harm activities into a state where there is a marked increase of affinity and engagement in digital self-harm behaviors and actions. This lens offers insight into how individuals develop a distorted sense of “normal” or “healthy” as they increase their
digital-self harm exposure. SCT’s construct of self-efficacy could help capture an individual’s confidence in engaging in self harm as well as an individual’s confidence in disengaging from self-harm behaviors based, in part, on content experienced online.

Cultivation Theory
Cultivation Theory (CT) is a social communications theory that describes the psychosocial impacts of media use. It was originally developed to better understand the effects of television viewing, yet is used to describe impacts of various other media platforms [41]. CT focuses on explaining how “repetitive, stable and persuasive” media inputs impact an individual’s perception [22,41]. Rather than directly causing specific behaviors, in CT, media exert a kind of “gravitational force” that pulls an individual deeper into the medium [41].

Research on cultivation theory has also found ties to psychosocial health, finding that media can influence beliefs, values and attitudes with impacts on loneliness and depression [26]. Because of the psychosocial nature of several of the co-morbid psychosocial disorders associated with digital self-harm, CT serves as a link to allow researchers to explore the connections between media influences and behavioral health. One question that CT could address is understanding the influences of “thinspiration” videos and images. Are extreme images more powerful or counterproductive? Is pervasive access the most influential feature of thinspo media?

Other Theoretical Constructs
There are numerous other frameworks that could potentially be used to unlock specific attributes of digital self-harm activities. The Transactional View of Self-Injury Behavior posits that distinct interactions exist between the physical, biochemical, social, and interpersonal environments [52]. Because it delineates health-seeking activity from both the act of promotion as well as those that are seeking, it could be useful when trying to understand both the seeker and the disseminator of destructive behavior signals within a digital context. The Problematic Internet Usage (PIU) Model connects the influence of Internet tools and online immersion to emotional, cognitive and behavioral issues within one’s offline interactions [9]. If one wanted a diagnostic tool to screen for levels of digital self-harm in an individual, the work around PIU could be foundational in the development of such a tool.

The theories outlined above are a starting point for understanding digital self-harm behaviors through foundational lenses that are well known in our field. Now that we have provided a definition for digital self-harm and have begun to operationalize it within the context of the CSCW field, our focus shifts to understanding the different dimensions of an agenda that supports this nascent vein of research.

SETTING THE AGENDA
We define digital self-harm as online communication and activity that leads to, supports, or exacerbates, non-suicidal yet intentional harm or impairment of the physical wellbeing of an individual.

In our proposed agenda, we describe research challenges in measuring correlation and causation associated with digital self-harm and its offline counterpart, designing of online affordances and experiences that may foster digital self-harm, and developing potential interventions. We close by discussing the implications of this agenda for the CSCW community.

Past Research
Currently there is a small corpus of research characterizing self-harm content and behavior online. These studies have started to examine the types of content that is shared [6,48], how information is shared online [20], how self-harm communities respond to moderation [11], and the ethical issues surrounding these platforms [55]. These studies have begun to expose the complexity and scope of digital self-harm within computing contexts.

Popular social media platforms are an ideal host for these types of explorations. Over the course of a decade, the focus of self-harm research activities migrated from investigating forums, message boards, and personal web spaces or blogs [20,55] to understanding the utilization of popular social media platforms like Instagram, Tumblr, and Twitter [11,48]. Pater et al. conducted a cross-platform analysis, including those listed above, of ED-related information shared online. This work developed a corpus of ED-related popular hashtags and a categorical taxonomy of online ED-related content [48]. Chancellor et al. developed an illness severity index using Instagram data to train a classifier that can analyze post content to denote the severity of an individual’s eating disorder [10]. Both Pater et al. and Chancellor et al.’s studies utilize the use of hashtags within social media to uncover community activities to an extent that studies like Shade’s analysis of weborexics [55] couldn’t because of the maturity and sophistication of the computing tools available at the time.

Research has also focused on behaviors related to the cutting community. In Moyer et al.’s exploration of self-injury behaviors online, they found that pro-cutting communities are especially worrisome – the community norms and the promoted materials overtly glorified destructive self-harm behaviors and employed language to actively encourage direct self-harm to individuals [42]. One example is that pro-cutting communities promote “triggering content.” They define this content as “online material that causes viewers to experience emotions and perceptions that typically precede and trigger cutting behavior.” [42]. Image 2 highlights this type of triggering content for cutters while Image 1 highlights this type of content for those dealing with eating disorders. [48].
This body of research begins to characterize self-harm activities presented online. These studies outline the types of activities that are found on social media platforms and websites. This research is limited in its explanatory power because it neither connects the online activities to individuals that have a clinical diagnosis, nor does it dive into the online presentations and their influence or impact on others. While these initial research contributions are fundamental to understanding digital self-harm at a high-level, a broader research agenda should encompass understanding the design affordances associated with online interactions and the impact of these experiences in promoting, facilitating, triggering, and sustaining self-harm behaviors.

This research agenda is flush with unanswered questions. How to identify the numerous forms of self-harm within digital contexts? Can online spaces be designed in a way that is more responsive to presentations of self-harm? What are the ethical responsibilities of online communities as it relates to individual expression of self-harm and the health impacts these activities have on the community members? Research is needed to understand how scope, entrenchment, degree of severity, and other factors of self-harm are exacerbated through digital exposure. Based on our survey of relevant literature, including the clinical definition of self-harm, an initial description of some digital self-harm online activities, and a distillation of related work; we outline common digital self-harm activities that warrant further research including:

- Propagating and proselytizing “norms” that assert self-harm as a legitimate alternative lifestyle through text and media.
- Providing tools and interactive features that allow one to track self-harm activities and measure against individual and shared goals.
- Curating content to directly trigger self-harm actions.
- Providing tips for how to hide self-harm actions from offline peers and family members.
- Recruiting and motivating individuals to engage in self-harm actions.
- Recruiting individuals to participate in online community activities that increase the commitment to self-harm.
- Sending reminders or other forms of “push” media to reinforce self-harm.

We organize our agenda through the perspectives of these research stakeholders: collaboration with clinical partners to gain a deeper understanding of the relationships between digital and offline self-harm behaviors; work by designers to understand the design affordances and online experiences of people consuming and producing digital self-harm content; work by platform owners and others seeking to create interventions online; and broader questions posed to the CSCW community.

Characterization, Correlation, and Causation

Open challenges for future research include fully characterizing the expression and promotion of self-harm activities online, advancing our understanding of correlations between digital activities and offline behaviors, understanding causal mechanisms, and identifying needed attributes for the development of interventions to address digital self-harm behaviors and activities.

In establishing the definition of digital self-harm, we highlighted the recent literature that has begun to characterize the landscape of digital self-harm activities. This critical first step is key to understanding the depth and reach of this health issue. This research has begun to characterize the terminology used in digital self-harm content and the types of media forms commonly used across a number of platforms like Instagram, Tumblr, and Twitter [11,48].

Questions regarding media include understanding how and why content migrates from one platform to another and understanding the level of influence difference types of media exert. With respect to understanding the influence of these communities on the individual exhibiting digital self-harm behaviors, both Cultivation Theory and Social Cognitive Theory can inform tools to analyze this issue. Cultivation Theory provides an understanding for the actual propagation of this content while Social Cognitive Theory provides explanatory power in relation to understanding patterns of behavior and behaviors at the scale of a community – not just the individual.

Key components central to this new research area extend to uncovering correlations and causations between increased online exposure and the physical manifestations of these disorders. Establishing correlations may be particularly helpful in screening Internet usage to identify afflicted individuals or individuals at high risk for self-harm behaviors. Employing data mining techniques within the context of SCT could facilitate new detection algorithms to identify those at high-risk. Current clinical self-harm screening tools do not reference online activity despite the growing prevalence of digital self-harm content and forums [48].

Multidisciplinary collaborations between clinicians and researchers will be critical in understanding the dynamic relationship between online and offline influences and presentations of the diseases where self-harm is a central characteristic. Some answers will come from theory. Theoretical constructs such as Social Cognitive Theory can give us a toolkit to understand and articulate the effects of environmental and social pressures on an individual. Other answers will come only through analyzing existing behaviors and practices. For example, recent research demonstrates the potential to identify signs of post-partum depression in social media postings, and, even ways to identify individuals at risk for post-partum depression following childbirth [13]. The analysis of social media...
postings by adolescents could identify those at risk for self-harm, potentially addressing critical health issues before they exact years of damage on an individual.

Understanding causation between online interactions and offline behavior will be challenging, but important for creating effective interventions. Exploring triggering content may be a productive first step. The use of SCT could be critical in understanding the path from first observing or engaging in digital self-harm activities and the physical manifestations associated with self-harm behaviors. Investigating the influence of different media types and attributes could help hone future Internet monitors. A critical challenge, that we will return to later in the paper, focuses on understanding how to support someone in recovery who nevertheless is still going to be online for daily life activities.

Design Features and Affordances
As highlighted earlier in this paper, recent research has focused on the familiarization and socialization of digital self-harm behaviors and activities taking place in social online spaces. Media content sites (e.g. Tumblr, Instagram, YouTube) and personal blogs have been the primary engagement zone for digital self-harm related to eating disorders and cutting. Through a deeper, and more contextual understanding of how digital self-harm is operationalized in online spaces, how do designers then demonstrate more sensitive and responsive design within these online spaces?

These communities look and feel different than more common or familiar communities, potentially because of the shocking and, in some cases, offensive content that is presented. This reaction might lead us to believe different design principles are at play, as compared to design principles traditionally associated with online communities. A deeper understanding of the drivers and design characteristics that afford these shocking presentations is needed. Studies examining deviant behavior communities online show that it is not the content, but the context and inter-dependent nature of the design decisions that create a “different” form of online third place that challenges these common practices [50].

The design features that can promote positive behaviors in other contexts can also be used to incite and encourage digital self-harm in others. How could designers address the challenges and discourage self-harm behaviors without disrupting positive aspects of certain design features? For example, the popular platform YikYak recently began allowing images to be shared through the platform. The moderation of images states “no inappropriate photos, illegal content, or faces will be allowed in local feeds”3 – a policy decision that aims to preserve the anonymity of the initial platform design. Research regarding the online media presentations of self-harm online highlight that individuals with eating disorders or those who self-injure through cutting often show body images that do not incorporate faces – they highlight specific body parts or aspects like collarbones, thigh gaps, and arms [20,48]. While trying to protect the majority of their users from inappropriate content and maintain anonymity within their community through design and policy, Yik Yak has made decisions that could prove to make the platform a fertile ground for digital self-harm activities because it supports the type of media popular within this domain.

For some forms of self-harm and other types of behavioral disorders, there are explicit environmental determinants that factor into the direct instantiation of the disease. Research shows that these factors include, but are not limited to, levels of aloneness, the amount of attention focused on the individual, the amount of engagement present, demands made on the individual, and access to preferable content [58]. Understanding these factors can aid in the development of more responsible design – taking into consideration screening and predictive models geared toward identifying individuals at risk for self-harm. Should designers take these determinants, or factors like them, into consideration during the design of digital spaces? If they do not, what does that mean for the potential negative outcomes that the evolving nature of digital self-harm might bring to bear within these online spaces?

Designing to mitigate or prevent digital self-harm may also require new best practices. For example, in the addiction recovery community, research has shown that anonymity and identity are inherent design tensions. Deciding how much anonymity to design into social computing systems may be critical in designing technologies and interventions that are sensitive to the needs of vulnerable populations [67]. Another tension that is critical in designing for digital self-harm is the struggle with balancing the undue influence of an overtly vocal minority with the general population of the platform [37]. Understanding how to appropriately reflect this balance this can be challenging, and extreme imbalances can lead to unintended outcomes within the community.

Behavioral Interventions
Examples of digital self-harm behaviors demonstrate how online spaces can be manipulated and transformed into platforms that promote, support, and exacerbate destructive behaviors. How can motivated designers and clinicians design interventions that, through online activity, promote healthy behavior both on- and offline?

Some online communities have taken initial steps in altering their site design and policies in recognition of these problems. Image 3 depicts the public service announcement that the Tumblr platform displays to a user if a predetermined self-harm term is used within their search

[^3]: http://blog.yikyakapp.com/post/128274737951/introducing-photos
Digital self-harm presents a unique problem - how to design interventions that result in lasting and persistent behavior change when the afflicted individual cannot be fully removed from the environment that causes or exacerbates the destructive behavior. Imagine you are a smoker who wants to stop, but you live in a world where smoking is allowed and most of your peer group smokes regularly, most of your social activity takes place in a tobacco shop, or your job requires you to work in a tobacco field? What is the likelihood that you will be successful in your goal of quitting smoking in this environment? The lack of social support as demonstrated through the presence of smokers within an individual environment will have substantial negative impacts on abstinence once the smoker has quit and is trying to maintain his smoke-free existence [38]. It is unreasonable to think that an individual being treated for self-harm could or would avoid all digital interactions. Understanding these constraints, is it possible for technological interventions to encourage lasting and meaningful change? These interventions won’t take place in a vacuum and will require an environment that actively promotes health and recovery.

Previous work in behavioral therapy highlighted the self-harm risks that individuals are exposed to when online [39]. A problem with these studies is that they compare deliberate measures of self-harm with internet behaviors such as viewing pornography, aggressive behavior and sharing identifying information online [39]. These studies do not specifically investigate digital self-harm behaviors. Overall, there is a gap for the behavioral therapy and the CSCW communities to work together to develop better in-situ behavioral treatments for patients as they begin their path of recovery from these mental illnesses.

Still further answers may come from studying other psychosocial therapy systems. Researchers studying online behavioral issues like cyberbullying and Internet overuse, may have useful lessons for digital self-harm. Mitigation tools have focused on preventing cyberbullying based on context-based filtering [34] and creating tools that bolster an individual’s self-esteem through asking friends of users to help send affirmations [62]. Recent research has utilized participatory design sessions with teens to distill design and implementation best practices for cyberbullying prevention and mitigation systems [4]. Is it possible to take successful clinical intervention strategies from other communities and find novel ways to integrate them into digital self-harm intervention strategies?

**Constant Connectivity and the Future of Online Interaction**

A blended identity or blended life that seamlessly integrates online and offline activities, a longtime vision for ubiquitous computing [14], is becoming more and more possible. In a world of constant connectivity and diverse online content, it is also now much easier to bias this blended existence to the extreme edges of the Internet. These extremes are visible in international and national politics, journalism, and the increased misogyny and hate-speech found in many online forums.

Should it be a goal for our field to make the Internet a “better” place? If so, who is to decide what is “better” and for whom it is “better”? There are examples from our field of where wanting to make the Internet a better place is a driving force for development, ICTD for example. There are examples of industry taking steps to curtail negative content through moderation, as discussed earlier in the paper. It is possible to understand how interaction online can foster activities that amplify the harm and impair the physical wellbeing of individuals. Should our community then also strive to understand how the Internet discourages – be it through design, practice, policy, or social compact – these interactions? This issue is contentious because it will require cultural debate on the boundaries between what is free speech and what is potential harm, which recent literature shows is a difficult prospect, regardless of platform [47]. If this stance holds, how can our community guide this process without infringing on the right for individuals to express free speech? While this paper doesn’t directly discuss the idea of the Internet becoming a “better” place, it does argue for ways that it could become a more informed, “safer” place with respect to the promotion of digital self-harm, with an ultimate aim of stepping up to the challenge of addressing this public health risk [28].
If we as a community agree that online activities that foster harm or directly impair the physical wellbeing of an individual do not make the Internet a better place, and that we as a community should do something about this type of activity, what are our first steps? One approach might be to promote a better offline and online balance, a concept that is not new to the CSCW field [54,63]. If this is the case, how should our community be designing and modeling appropriate behavior of this offline and online balance? One approach is designing online interactions that require or motivate offline activities, such as electronic games that require physical activity or monitors that alert an individual to excessive levels of screen time [65]. These types of ideas are a great start, but there are many design approaches that have yet to be fully explored.

How can we imagine reshaping life online that is grounded in realistic moderation: conveying messages that most women do not look like supermodels, most kids do not cut themselves, and that everyone feels lousy and alone at times? As creators of this new pervasive, blended existence, what is the future role of the CSCW community as we stare into a future of our creation?

CONCLUSION

If the CSCW community represents the main stakeholders designing, building, and researching online spaces, then we have a responsibility to understand how these advancements are used, whether that is for positive individual and societal benefit or the negative. By not studying digital self-harm, the community will continue its prosocial bend of only looking at positive, and not the negative, impacts of the tools and platforms that we develop. Public health data shows that this mental illness will only continue to increase in pervasiveness [15]. The multi-disciplinary approach needed to make sense of digital self-harm, how it manifests and develops, and how it can be addressed through responsive platforms and interventions is a non-trivial undertaking of great importance. As the community begins to answer the outlined questions in this research agenda, we will begin to gain deeper insights into the connections between online spaces and mental health. Because this research will catalyze multi-disciplinary collaboration, the ability to translate these connections into meaningful tools and treatments is a tangible and likely outcome. Defining digital self-harm is an important first step.

ACKNOWLEDGEMENTS

We are grateful for the feedback from our colleagues in the Everyday Computing Lab (Andrew D. Miller, Maia Jacobs, James Clawson) and the reviewers. We would also like to thank Nazanin Andalibi, Oliver Haimson, Stevie Chancellor and Munmun De Choudhury for their collaborations in this space. Additionally, we are thankful to the students and faculty at Georgia Tech for the community support given throughout this research. This research was supported in part by NSF award SHB-1116801.

REFERENCES

12. Christopher AD Charles. 2014. Skin Bleaching,


37. Lena Mamykina, Bella Manoim, Manas Mittal,


63. Sherry Turkle. 2012. *Alone Together: Why we expect more from technology and less from each other*. Basic Books.


