Social Safety Theory: Conceptual foundation, underlying mechanisms, and future directions

George M. Slavich a, Lydia G. Roos a, Summer Mengelkoch a, Christian A. Webb b,c, Eric C. Shattuck d, Daniel P. Moriarity a and Jenna C. Alley a

aDepartment of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA; bMcLean Hospital, Belmont, MA, USA; cHarvard Medical School, Boston, MA, USA; dInstitute for Health Disparities Research and Department of Public Health, University of Texas at San Antonio, San Antonio, TX, USA

ABSTRACT
Classic theories of stress and health are largely based on assumptions regarding how different psychosocial stressors influence biological processes that, in turn, affect human health and behavior. Although theoretically rich, this work has yielded little consensus and led to numerous conceptual, measurement, and reproducibility issues. Social Safety Theory aims to address these issues by using the primary goal and regulatory logic of the human brain and immune system as the basis for specifying the social-environmental situations to which these systems should respond most strongly to maximize reproductive success and survival. This analysis gave rise to the integrated, multi-level formulation described herein, which transforms thinking about stress biology and provides a biologically based, evolutionary account for how and why experiences of social safety and social threat are strongly related to health, well-being, aging, and longevity. In doing so, the theory advances a testable framework for investigating the biopsychosocial roots of health disparities as well as how health-relevant biopsychosocial processes crystalize over time and how perceptions of the social environment interact with childhood microbial environment, birth cohort, culture, air pollution, genetics, sleep, diet, personality, and self-harm to affect health. The theory also highlights several interventions for reducing social threat and promoting resilience.

ARTICLE HISTORY
Received 14 June 2022
Accepted 19 January 2023

KEYWORDS
Stress; Resilience; Development; Biology; Health; Disease

Introduction
One of the most consistent findings in the social and health sciences involves the sizable impact that both positive and negative social factors have on human health, longevity, and behavior. Whereas a large literature has shown that social support predicts a variety of positive health outcomes, including reductions in both disease-specific and all-cause mortality, there is also substantial evidence that social stressors such as abuse and neglect negatively impact the development and function of most major organ systems that affect health (Freak-Poli et al., 2021; Lupien et al., 2009; O’Connor et al., 2021; Wulsin et al., 2022). Given the tremendous amount of research that has been done investigating associations between stress and health, one might think there is consensus regarding which social factors are most beneficial and deleterious for human health, but this is not the case (Cohen et al., 2019; Simmons et al., 2021). In fact, there is little agreement on this topic (Epel...
et al., 2018), leading some to suggest that researchers should avoid using the term ‘stress’ altogether (e.g., Kagan, 2016). Yet others have dealt with these complex issues by avoiding them altogether, such as those in the tongue-in-cheek, but all-too-real field of stressnology (Slavich, 2019).

How did we get here? One explanation is that we have historically built theories of stress, coping, and resilience based on assumptions regarding how we think social experiences might affect health-relevant biology. This strategy amounts to armchair theorizing that yields predictions that may or may not be well-aligned with the evolved regulatory logic or functioning of the underlying systems that affect disease risk. This scientific guessing has occurred for a few reasons, but a main cause is that tools for assessing neural, physiological, immune, and genomic responses to social situations have only recently become affordable and more widely available (Slavich, 2020a). The second reason for our relatively imprecise understanding of how positive and negative social experiences affect health centers on the fact that this research has generally been confirmatory in nature with few attempts to disprove the theories or investigate their specificity. As an example, consider the finding that changes at work predict depressive symptoms (e.g., Li et al., 2013). This is not surprising, given that change can be stressful. However, because work changes can result from many different circumstances – everything from an elected occupational change to unwanted social demotion – we must also ask: is degree of change more predictive than other aspects of this situation, such as the degree of social devaluation or rejection experienced? By focusing on change alone, we will conclude that change is indeed relevant, but this effect can be a red herring that distracts us from examining other stressor characteristics that may matter more.

To address these issues, we propose that associations between social experiences, health, and behavior may be better understood through the lens of evolutionarily grounded frameworks such as Social Safety Theory. Rather than defining the key characteristics of positive and negative social experiences based on biological guesses or behavioral data alone, Social Safety Theory aims to use knowledge regarding the primary goal and regulatory logic of the human brain and immune system as the basis for predicting which social situations and experiences these health-relevant biological systems should be most sensitive and responsive to in order to maximize reproductive success and survival. Going from biology to psychology in this way is not foolproof as it can still lead investigators astray, but we believe it represents a more fruitful approach, as it permits researchers to conceptualize positive and negative constructs, such as social support and social adversity, in ways that are biologically grounded and evolutionarily informed, thus refining our fundamental understanding of stress biology and health.

In the present article, we extend work on Social Safety Theory by first summarizing the theory, defining its key constructs, and explaining its evolutionary basis. Second, we describe the main biological and social-cognitive processes underlying the theory, and examine how culture and social institutions affect experiences of social safety and threat. Third, we discuss how Social Safety Theory can be applied to study health disparities in vulnerable groups and compare the theory to several related formulations of stress and health. Fourth, we examine nine key factors that can impact the social signal transduction pathways underlying the theory – namely, childhood microbial environment, birth cohort, culture, and air pollution (i.e., situational factors), and genetics, sleep, diet, personality, and self-harm (i.e., individual difference factors). Fifth, we specify construct measurement and theory testing issues that should be taken into account in future research on this topic. Finally, we highlight several interventions for reducing social threat and fostering social safety at the individual and collective level. In covering these topics, we aim to identify how Social Safety Theory may help address critical questions in human health and resilience, and what topics would benefit from additional investigation to advance our understanding of how positive and negative social factors affect health, well-being, aging, and longevity.

Social Safety Theory: A biologically based evolutionary perspective on life stress, health, and behavior

Social Safety Theory aims to account for how and why specific types of positive and negative social experiences are strongly related to human health and behavior (Slavich, 2020a, 2022). The theory
was developed to help advance classic thinking on this topic, which has persisted until today and driven an overly general, sometimes misguided approach to investigating the specific types of experiences hypothesized to be most strongly associated with health-related outcomes. In the case of adverse life experiences, for example, Selye (1976) argued that stress is ‘the nonspecific response of the body to any demand’ (p. 74) and that a stressor is ‘that which produces stress’ (p. 78). Guided by this nonspecific view of life stressors and stress physiology, most studies conducted today still use stressor exposure metrics that boil a person’s myriad of stressful experiences down to one total score that completely ignores the specific stressors that occurred and when exactly they happened, thus precluding an examination of whether some stressors are more predictive than others (Monroe & Slavich, 2020; Slavich, 2016, 2019).

Several more sophisticated frameworks have since been proposed. For example, Holmes and Rahe (1967) posited that a stressor’s impact is related to the degree of change or upheaval the experience typically causes, Lazarus and Folkman (1984) argued that stress arises when situational demands exceed an individual’s ability to adequately cope, and Clark and Beck (1999) theorized that stressors can be sorted into different life domains such as interpersonal (e.g., intimate relationships) and achievement (e.g., work), and that a stressor’s impact is heightened when its content matches the person’s cognitive vulnerability. Although plausible at the psychological level, the utility of these frameworks has been limited, as these distinctions have not consistently predicted health outcomes. What has been missing, it seems, is a grounding of these types of distinctions in a solid understanding of the biological processes that are most directly implicated in disease risk, and the social-psychological experiences that should most strongly activate and sustain these health-damaging processes given the basic function and regulatory dynamics of the underlying systems.

A Highly Evolved, Integrated, Prediction and Anticipatory Response System

Social Safety Theory aims to address these theoretical limitations by viewing the relevance of positive and negative social experiences from the standpoint of the human immune system and brain. As described by Slavich (2020a, 2022), animal and human brains evolved over millennia to respond to social-environmental cues in a way that maximizes the likelihood of reproduction and survival. This function is largely accomplished through allostatic, or the processes of anticipating needs and marshalling social, biological, and behavioral resources to maintain homeostasis in the face of a continually changing, sometimes threatening environment (McEwen & Wingfield, 2003; Sterling, 2012). Although the brain is the primary organ responsible for allostatic, it relies on multiple sensory mechanisms, one of which is the immune system – an incredibly complex and fascinating system, frequently overlooked in neuroscience research, which continually monitors the internal milieu for cues indicating biological danger including tissue damage and infection (Slavich, 2020b). Given the extensive breadth and depth of bidirectional connections between the immune system and brain, the immune system has been theorized as a ‘sixth sense’ that shares inter – and intra-system chemical messengers (e.g., peptide and nonpeptide neurotransmitters, cytokines) with the brain that enable continual bidirectional communication and calibration (Blalock, 2005). In addition to recognizing conserved features of microbes such as bacterial and viral genomes, the immune system can detect cellular stress or death caused by tissue damage, bodily trauma, and ischemia (Kawai & Akira, 2006; Meizlish et al., 2021), and modulate its functional activity and capacity based on past pathogen encounters (Mayer et al., 2019) and the surrounding microbial environment (Rook et al., 2017).

One of the most notable features of the immune system for stress research is its ability not only to react to bodily damage or danger after it has occurred but to mount anticipatory responses to threats that may occur based on past microbial exposures and social-environmental experiences (Slavich, 2020a, 2022). This ability confers significant adaptive advantage, as preemptive activation of the immune system mobilizes immune cells to sites of potential injury or infection, which accelerates wound healing and recovery following tissue damage, and thus improves the odds of reproduction and survival (Schiller et al., 2021). To accomplish this task, the immune system relies on the brain to
Social Safety Theory is grounded in the understanding that the primary purpose of the human brain and immune system is to keep the body biologically and physically safe. To accomplish this challenging task, humans developed a fundamental drive to create and maintain friendly social bonds and to mount anticipatory biobehavioral responses to social, physical, and microbial threats that increased risk for physical injury and infection over the course of evolution. (a) Accordingly, the brain continually monitors the (1) social environment, interprets social signals and behaviors, and judges the extent to which its surroundings are socially safe versus threatening. These appraisals are subserved by the (2) amygdala network, mentalizing network, empathy network, and mirror neuron system (i.e., the social brain). When a potential social threat is perceived, the brain activates a multi-level response that is mediated by several social signal transduction pathways — namely, the (3) SNS, (4) HPA axis, (5) vagus nerve, and (6) meningeal lymphatic vessels. These pathways enable the brain to communicate with the peripheral immune system and vice versa. Whereas the main end products of the SNS (i.e., epinephrine and norepinephrine) suppress transcription of antiviral type I interferon genes (e.g., IFNA, IFNB) and upregulate transcription of proinflammatory immune response genes (e.g., IL1B, IL6, TNF), the main end product of the HPA axis (i.e., cortisol) generally reduces both antiviral and inflammatory gene expression but also can lead to increased inflammatory gene expression under certain physiologic circumstances (e.g., glucocorticoid insensitivity/resistance). The vagus nerve, in turn, plays a putative role in suppressing inflammatory activity, whereas meningeal lymphatic vessels enable immune mediators originating in the CNS to traffic to the periphery, where they can exert systemic effects. (b) This multi-level ’Biobehavioral Response to Social Threat’ is critical for promoting well-being and survival. However, it can also increase a person’s risk for negative health and behavioral outcomes when it is sustained by internal physiologic or external social recursion. Several factors can also moderate these effects, including birth cohort, culture, personality, childhood microbial environment, sleep, genetics, air pollution, diet, and self-harm. A person’s developmentally derived social safety schemas play a particularly important role in this multi-level response as they shape how social-environmental...
detect increased risk for physical danger as far in advance as possible. In addition to enabling the preemptive mobilization of immune system resources to threat, this highly sophisticated neural-immune integration maximizes a person’s ability to successfully deal with a variety of biobehavioral challenges by modulating energy, metabolism, blood pressure, temperature, diet, and sleep to meet the demands of the situation (Schiller et al., 2021). A complete account of brain regions involved is available elsewhere (e.g., Kraynak et al., 2018) but, in short, includes the amygdala network, mentalizing network, empathy network, and mirror neuron system among others (see Slavich, 2020a). As depicted in Figure 1, these systems in turn regulate peripheral processes – and vice versa – via the sympathetic nervous system (SNS), hypothalamic–pituitary–adrenal (HPA) axis, vagus nerve, and meningeal lymphatic vessels (see Slavich, 2020a, 2022).

**The Critical Relevance of Safety and Threat**

If we pause for a moment and think about the types of social-environmental circumstances that are most likely to engage this highly integrated, multi-level threat response, current physical danger would be atop the list, as it represents the most urgent threat to reproductive success and survival. Among social experiences, though, given the abovementioned advantages of mounting preemptive biobehavioral responses to threat, Social Safety Theory posits that the Biobehavioral Response to Social Threat depicted in Figure 1 should be most strongly activated by social circumstances that indicated an increased risk of physical danger over the course of evolution, such as those involving social conflict, aggression, devaluation, discrimination, isolation, rejection, or exclusion (Slavich, 2020a, 2022). Consistent with this possibility, a large literature has now accumulated showing that stressors involving these characteristics are consistently – and oftentimes most strongly – related to increases in inflammatory activity, which can promote inflammation-related disease conditions (e.g., depression, heart disease, neurodegenerative disorders), and with changes in both cellular and humoral immunity, which affect the ability to successfully degrade viruses and bacteria (Murphy et al., 2015; Murphy et al., 2013; Segerstrom & Miller, 2004; Slavich & Irwin, 2014; Steptoe et al., 2007). These circumstances also involve differing degrees of psychological upheaval and change, as well as many other stressor characteristics. What Social Safety Theory posits matters most for predicting health-related outcomes, though, is the extent to which a situation is appraised as lacking social safety and possessing social threat, which we define in detail below.

**Key Tenets of Social Safety Theory**

As described by Slavich (2020a, 2022), this primary goal and regulatory logic of the human brain and immune system translates into three main tenets of Social Safety Theory, which advance our understanding of stress biology and elucidate the biological bases of human health and behavior. The three tenets are:

- **Tenet 1:** Humans evolved to foster social safety – namely, humans exhibit a fundamental drive to develop and maintain friendly social bonds, especially with in-group members.

- **Tenet 2:** Social safety is beneficial for human health and behavior – namely, social-environmental experiences indicating safety have broadband benefits for a variety of health and behavioral outcomes, including longevity.

- **Tenet 3:** Social threat is harmful to human health and behavior – namely, social-environmental experiences indicating threat are associated with a variety of negative health and behavioral outcomes, including mortality.
Evidence supporting these tenets comes from a variety of fields and broadly supports the fundamental notion that experiences of ‘social safety and social threat lie at the heart of life’s most impactful experiences’ (Slavich, 2020a, p. 287). To better understand how these tenets can be tested, we turn next to describing the key constructs of Social Safety Theory.

**Key constructs of Social Safety Theory**

Social Safety Theory encompasses three key constructs that influence the extent to which social experiences promote vs. degrade health. These three constructs are *socially safe and socially threatening situations*, *perceived social safety and social threat*, and *social safety schemas*. Across each construct are two distinct, non-mutually exclusive (i.e., orthogonal) concepts – social safety and social threat – that can be measured by the actual social-psychological characteristics of a situation as well as by a person’s perceptions or appraisal of the situation.

**Socially Safe and Socially Threatening Situations**

As described by Slavich (2020a), *socially safe situations* are social circumstances characterized by social acceptance, understanding, inclusion, connection, belonging, cohesion, harmony, support, validation, predictability, stability, and authenticity that would have conferred evolutionary benefits. Conversely, *socially threatening situations* are social circumstances characterized by social conflict, aggression, devaluation, criticism, disapproval, discrimination, isolation, rejection, exclusion, turbulence, unpredictability, manipulation, and betrayal that would have conferred evolutionary costs. Situational characteristics of social safety and social threat thus represent the observable features of a particular life event or social situation.

To be considered *social*, a core feature of the experience must involve an interpersonal exchange or have direct interpersonal implications. Given that sociality imbues nearly every aspect of human life, life events and situations that could generally be considered nonsocial in terms of their primary characteristics may well initiate social stressors and have social repercussions, but these latter circumstances should be investigated separately. For example, receiving a poor score on an exam can result in negative social evaluation and criticism by peers or family members, and a job loss – although economic in nature – may precipitate a loss of close social ties at work or social status among friends and family members. Conversely, not all social effects must be negatively valenced just because the initial nonsocial stressor was negative (and vice versa for positive experiences). For example, receiving a poor exam score can provide an opportunity to commiserate and seek support from – and provide support to – others who received a similar score. Therefore, although these experiences may be bound together, they should be assessed as separate, albeit linked situations to ensure precise measurement.

**Perceived Social Safety and Social Threat**

*Perceived social safety and social threat* represent an individual’s appraisal of a particular social situation or interaction as conferring social safety and threat, which is obviously influenced by, but still separable from, the observable characteristics of situation. For example, an individual may be party to an ambiguous situation (e.g., good friend walks by and does not say hello) but interpret that situation as socially rejecting. In social instances such as this, *perceived social safety* occurs when a person perceives that they are socially accepted, understood, connected, cohesive, harmonious, supported, validated, or belong, and that others and the person’s relationships with those individuals are predictable and stable. Conversely, *perceived social threat* occurs when a person perceives social conflict, aggression, devaluation, criticism, disapproval, discrimination, isolation, rejection, exclusion, manipulation, or betrayal from others, or that others and the person’s relationships with those individuals are unpredictable or otherwise turbulent.
The distinction between socially safe and socially threatening situations vs. perceived social safety and social threat involves the critical difference between (for example) the extent to which someone feels personally rejected after getting broken up vs. the consensus judgment that the individual is experiencing a life event characterized by social rejection and is thus a socially threatening situation. Clearly, the observable characteristics of a given social situation constrain the cognitive appraisals that are most likely to be generated insofar as someone who gets broken up with is much more likely to feel socially rejected than included in that moment. Likewise, someone who has been broken up with is more likely to feel socially rejected than someone who, for example, scored poorly on an exam. However, the extent to which a person appraises a breakup (or any other interpersonal stressor) as socially rejecting can vary, and these individual differences in appraisal are critical to measure in addition to the observable characteristics of the situation given that neurocognitive perceptions of social safety and social threat strongly influence the downstream biological consequences of these experiences and, therefore, the effect they have on health and behavior (Slavich & Cole, 2013; see also Shields et al., 2023). As with stressor appraisals more generally, perceived social safety and threat are conceptualized as an acute experiential state resulting from a particular social situation or event (Lazarus & Folkman, 1984). In turn, a person’s propensity to perceive social situations as safe vs. threatening is thought to be informed by their more trait-like social safety schemas, described below. Examples of socially safe and socially threatening situations, and perceived social safety and social threat, are provided in Table 1.

**Social Safety Schemas**

Finally, social safety schemas are conceptualized as cognitive representations of the social self, world, and future, which – unlike other schema-focused frameworks – specifically involve the dimensions of social safety and social threat. In this context, social self-schemas concern an individual’s conception of their own characteristics, attributes, virtues, and resources, such as whether they belong, are lovable, deserve affection and respect from others, and have the ability to cope or get through challenging interpersonal situations. Social world schemas, in turn, reflect the extent to which an individual generally regards other people as friendly versus hostile, predictable versus unreliable, supportive versus critical, helpful versus hurtful, and sincere versus manipulative. Finally, social future schemas concern a person’s cognitive representation of their social future as involving social inclusion vs. isolation, stability vs. instability, interpersonal success vs. failure, and safety vs. danger. Together, these three types of schemas play critical roles in structuring social attitudes, expectations, beliefs, and behaviors across the life course. For example, they shape the thoughts and expectations one might have about their own tendencies, strengths, and weaknesses when it comes to interpersonal situations, as well as what an individual can expect from others, the social world, and the future. This framework thus includes more than just self-schemas, which are only one schema category in Social Safety Theory. The framework also goes beyond attachment theory schemas by systematically including representations of the social self, world, and future. Examples of socially safe and socially threatening beliefs, organized by the three main types of social safety schemas, are depicted in Figure 2.

As discussed by Slavich (2020a), social safety schemas are hypothesized to be shaped by the social experiences a person has over their lifetime, as well as by the messages they received about the self, others, and future, and by their interpretations of these situations and messages. Childhood and adolescence is thought to be a key developmental period for the formation of these schemas, as children’s brains are exquisitely sensitive to social input and are continually gauging the extent to which the social environment is safe vs. threatening (Somerville, 2013). To accomplish this task, the brain pays close attention to the actual situations a person encounters (e.g., being frequently hugged or included vs. physically harmed or rejected), but also to social and cultural norms, and to the messages received and meaning that caregivers and peers attribute to socially salient events. Sometimes the messages that peers and caregivers convey about the social world are
explicitly negative, such as ‘That person is dangerous,’ ‘Other people can’t be trusted,’ or ‘People only care about themselves,’ but even ambiguous, seemingly helpful messages can impact the developing mind’s construal of the social self, world, and future, such as ‘I’m not sure about him,’ ‘Just be careful around her,’ ‘Good luck finding love,’ or ‘Remember not to walk down that street on your way home.’ The intended goal here is undoubtedly to ensure safety, but the implicit message conveyed is one of threat.

These experiences and appraisals are, in turn, hypothesized to shape a person’s expectations, beliefs, and behaviors, which include their perceptions and appraisals of their social self, world,

Table 1. Examples of socially safe and socially threatening situations, and perceived social safety and social threat. Adapted from Diamond & Alley (2022) to reflect general population experiences.

<table>
<thead>
<tr>
<th>Socially Safe &amp; Socially Threatening Situations</th>
<th>Perceived Social Safety &amp; Social Threat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Safety</strong></td>
<td></td>
</tr>
<tr>
<td>Being able to interact with other individuals, in person or online, who share common interests</td>
<td>Believing you are around other people who understand you</td>
</tr>
<tr>
<td>Safe Schools Programs that explicitly condemn bullying: training teachers and staff to respond to such instances</td>
<td>Believing your school wants you to feel and be safe, and feeling that teachers and staff members are dedicated to supporting you</td>
</tr>
<tr>
<td>Seeing people (e.g., friends, peers, the media) and institutions condemn the mistreatment of others</td>
<td>Believing those around you think the mistreatment of others is wrong and will protect you if you needed</td>
</tr>
<tr>
<td>Having family members who provide affection and reassurance, ask about your life, and include, support, and protect you</td>
<td>Believing family members care about your friends and life, and will provide support and protection if needed</td>
</tr>
<tr>
<td>Having family members and friends welcome your friends and loved ones when they meet them</td>
<td>Believing your family and friends fully embrace you, your friends, and your loved ones</td>
</tr>
<tr>
<td>Encountering inclusive language on standard forms (e.g., being able to correctly describe your identity, race, ethnicity, gender, sex)</td>
<td>Believing people in official offices understand and respect diversity, and want to represent you accurately</td>
</tr>
<tr>
<td>Seeing family members, friends, and colleagues speak up right away to protect you, and that you do not have to defend yourself all alone</td>
<td>Believing people around you will step forward when you or someone else is being threatened, devalued, or harmed</td>
</tr>
<tr>
<td>Having people make room for you at a table, on a bus, in a car, etc.</td>
<td>Believing people are comfortable being physically near you</td>
</tr>
<tr>
<td>Having people ask you thoughtful questions about your life and check in with you before making assumptions about what might make you comfortable in social situations</td>
<td>Believing people are genuinely interested in your whole self, care about your comfort and safety, and respect your autonomy and judgement</td>
</tr>
<tr>
<td>Receiving positive comments, likes, and shares when you reveal meaningful information about yourself on social media or in person</td>
<td>Believing people – even those who do not know you personally – like, support, and appreciate you</td>
</tr>
<tr>
<td>Having people confide in you</td>
<td>Believing you are someone others can trust</td>
</tr>
<tr>
<td><strong>Social Threat</strong></td>
<td></td>
</tr>
<tr>
<td>Witnessing bullying at school or online</td>
<td>Fearing bullying at school or online</td>
</tr>
<tr>
<td>Witnessing violence or harassment</td>
<td>Fearing violence or harassment</td>
</tr>
<tr>
<td>Being rejected or excluded by friends or family members</td>
<td>Fearing the loss of social connection with friends or family members</td>
</tr>
<tr>
<td>Losing practical or emotional social support</td>
<td>Fearing the absence or loss of practical or emotional social support</td>
</tr>
<tr>
<td>Being governed by laws that are unequal, do not value your identity or beliefs, or restrict your access to health care, social, or other services</td>
<td>Fearing that your city, community, or country is biased and does not protect you or your access to key medical, social, or other services</td>
</tr>
<tr>
<td>Having family members block your access to peers or your community</td>
<td>Fearing the loss of social support from your peers or community</td>
</tr>
<tr>
<td>Reading religious doctrine or attending religious services that devalue, reject, or restrict the expression of your identity</td>
<td>Fearing for your spiritual salvation due to your identity, values, or beliefs</td>
</tr>
<tr>
<td>Hearing people make derogatory jokes or comments about you or groups to which you belong</td>
<td>Fearing that you will be devalued, mistreated, or excluded by people or groups you value</td>
</tr>
</tbody>
</table>
and future; how they navigate the social world, including the types of relationships they develop and how they act in those relationships; and how their brains and immune systems respond to social circumstances, which together impact other aspects of human cognition and biology, and – ultimately – health. As such, social safety schemas are thought to be both informed by and to influence the social situations a person experiences. Due to social safety schemas being especially malleable during sensitive periods of development, Social Safety Theory hypothesizes that these schemas are relatively trait-like; however, they may indeed change over the life course as a function of experiencing particularly impactful social life events, such as finding a highly dependable and loving romantic partner (positive) or experiencing a socially traumatic life event (negative).

Finally, Social Safety Theory posits that social safety and social threat are not mutually exclusive or on opposite ends of the same continuum. Therefore, although aspects of social safety and threat may certainly be experienced disparately, they may also be experienced simultaneously. Whereas destructive criticism (i.e., criticism that undermines and harms the recipient) entails the presence of social devaluation and reduced social safety, for example, constructive criticism involves social connection and support in the presence of social evaluation, and can thus be perceived as involving both social safety and threat. Similarly, a social relationship may be perceived as safe in some respects but threatening in others. For example, a person may provide unconditional acceptance to their friend while at the same time exhibiting unpredictable or undependable behavior, such as not keeping social commitments. Finally, as depicted in Figure 2, people can hold schemas that include beliefs of both social safety and threat, as shaped by their prior experiences with both types of situations.

Figure 2. Examples of socially safe and socially threatening beliefs, organized by the three main types of social safety schemas.
Comparison with other constructs

Although the notion that positive and negative social experiences strongly influence health and behavior has been described before (e.g., Ainsworth et al., 1978; Baumeister & Leary, 1995; Gilbert, 2005), our understanding of how different features of the social environment relate to biological stress processes in humans has been limited. Here, we contrast the constructs described above with related constructs to provide a clearer understanding of Social Safety Theory.

Social Stressors vs. Social Threat

“Social stressor” is a term that has been used very permissively to describe a wide variety of interpersonal experiences. Whereas such stressors can be regarded as socially challenging (e.g., ‘This audition is challenging, but presents an opportunity for me to shine’), they can also be appraised as socially threatening (e.g., ‘My teacher is trying to make me fail’). Social threat refines this broader category of social stressors to focus specifically on the latter – namely, interpersonal events and situations that are hypothesized to strongly activate the multi-level Biobehavioral Response to Social Threat depicted in Figure 1.

Identifying social threat can be tricky, as it requires intimate knowledge of what actually happened as well as how the individual perceived the situation. For example, a heated argument with a spouse may be an opportunity for explanation, acceptance, and relationship growth or a situation involving social conflict, criticism, and rejection. In these cases, the extent to which a social stressor involves social threat will depend on the specific details of what occurred as well as the individual’s perception of the situation, which Social Safety Theory hypothesize will be shaped by their social safety schemas. Likewise, perceived social threat can arise from situations that most people would not consider a social stressor. For example, an ambiguous comment such as ‘What happened to you?’ can be perceived as socially threatening by someone with a highly critical parent but as caring and supportive by someone else. Even situations that would generally be considered socially safe, such as asking a friend for a ride to the airport, can be perceived as socially threatening if the person asking for the ride believes the request puts them at a disadvantage or in debt in the relationship, or if he or she appraises the experience as indicating to others an inability to plan or take care of their own needs. In short, whereas ‘social stressors’ has been used to refer to both socially challenging and threatening circumstances, as well as those that include varying elements of social safety and threat, we define social threat as that which would have increased individuals’ risk for physical danger over the course of evolution.

Psychological Safety vs. Social Safety

Psychological safety is another term that should be clarified vis-à-vis social safety. In the organizational psychology literature, *psychological safety* is the shared belief held by members of a team that the work environment is safe for interpersonal risk-taking (Edmonson, 1999). That is, workers believe they can voice new ideas, experiment with novel approaches, collaborate with others (even if it results in failure), and willingly seek and provide honest feedback, all without fear of experiencing negative consequences such as shaming or ridicule that might result in embarrassment (Edmonson, 1999; Newman et al., 2017). Workplace psychological safety is often characterized by the perception that a social environment fosters acceptance, inclusion, belonging, and support, which is similar to perceived social safety in a group setting. As described by Slavich (2020a, 2022), however, social safety goes far beyond the workplace and describes not only a collective, team-level social climate of safety, but also the perception of social safety in all the ‘social circles,’ or interpersonal contexts, in which people interact, including family, neighborhood, city, and country. Moreover, social safety can be measured as a person’s experiences of safety as tied to a specific interpersonal life event or social situation, or as an individual’s broader social safety schemas, which are hypothesized to give rise to specific social cognitions and behaviors in daily life.
**Dimensions of Deprivation and Threat vs. Social Safety and Threat**

Another distinction that can be drawn is between the dimensions of deprivation and threat (by Sheridan & McLaughlin, 2014) vs. social safety and threat (by Slavich, 2020a). As described by Sheridan and McLaughlin (2014), childhood adversity can be characterized as involving deprivation, defined as the absence of expected environmental inputs and complexity, and threat, defined as the presence of threat to one’s physical integrity. Social Safety Theory’s conceptualization of social safety and threat differ from this perspective in a few key ways. First, whereas the deprivation and threat model focuses specifically on the potential for physical harm as being critical, Social Safety Theory focuses mainly on social cues that would have historically indicated an increased risk for physical harm, such as interpersonal conflict or hostility. Second, in contrast with the deprivation and threat model, Social Safety Theory accords distinct and separable roles for positive and negative social experiences in influencing health and behavior, with social safety and social threat being orthogonal dimensions of human experience that can be measured separately and likely have different effects on the brain and immune system. This conceptualization also highlights the presence of social safety as being a key, health-promoting feature of social-environmental circumstances involving low deprivation – something not underscored by the deprivation and threat model.

**Basic evolutionary dynamics and principles underlying Social Safety Theory**

To fully understand the central relevance of social safety and threat for stress biology, we must take a step back and consider how these systems evolved and why they operate as they do. That analysis reminds us that in order to survive, organisms must overcome numerous challenges – including physical threats from the environment, social and physical threats from dangerous others, and biological threats from pathogens – for at least long enough to find a mate, reproduce, and ensure that their offspring are healthy enough to do the same. Throughout our evolutionary history, humans are one of many species that evolved a complex social dynamic to meet these demands. Living as a social organism required humans to navigate social relationships in ways that fostered social cooperation and, when possible, avoided social conflicts. Ancestors of modern humans lived in social groups for millions of years and evolved complex psychological, biological, and behavioral capabilities that helped them maximize the benefits and minimize the costs of sociality (Henrich, 2015). Consequently, modern humans now possess empathy (MacLean, 1990) alongside ‘cheater detection’ mechanisms (Cosmides, 1989), in-group preferences (Bloom, 2013) alongside aggression (Choi & Bowles, 2007), and prosocial tendencies (Henrich, 2015) alongside behavioral immune systems (Schaller & Park, 2011), which have combined to shape our evolutionary trajectory.

Living within complex social systems has many fitness benefits but also involves serious challenges. In terms of benefits, living in groups enabled humans to make large advancements in living conditions, health, and longevity by enabling people to work together to more safely hunt larger game, protect each other from hostile neighboring groups, and specialize tasks leading to tool innovation. Years of research have found that increased social integration substantially decreased mortality risk in both humans (Holt-Lunstad et al., 2010; Vila, 2021) and non-human animals such as baboons (Archie et al., 2014; Silk et al., 2010). For example, baboon infants born to mothers who are highly socially integrated are more likely to survive into their second year of life as compared to those born to mothers who are less socially integrated (Silk et al., 2003). These benefits of social integration for a member of a social species also make obvious the costs of social threats such as ostracization. Just as social integration decreases mortality risk, being socially ostracized greatly increases the risk of death from starvation, predation, and infection across all social species (Williams, 2007). Indeed, ostracism in ancestral human populations has been described as ‘social death’ due to the large impact that cutting social connections has on survivorship and the ability to successfully reproduce (Williams, 2007).
Due to strong selection pressures for maintaining a robust and reliable social network – in other words, maintaining social safety – social animals use social exclusion as a tool to enforce social norms, which can benefit the group members by minimizing intra-group conflict and reducing disease risk (Kurzban & Leary, 2001). Oftentimes, this is carried out through the maintenance of status hierarchies and in territorial disputes across the animal kingdom (Wilson, 2000). For example, chimpanzees who do not submit to higher-status group members are often forced out of the group to prevent further conflict within the group (Goodall, 1986). Many social species, from humans to fish, also socially isolate or avoid members that give off illness cues (Dugatkin et al., 1994; Schaller, 2011), presumably to protect oneself and the group from the potentially deadly threat of disease.

Humans, and some other social animals, are thus theorized to have evolved a repertoire of adaptive psychological, behavioral, and biological responses to navigate social threats. For example, humans are theorized to possess psychological mechanisms that accurately detect threats to social inclusion and motivate behavioral responses that increase social safety, such as responding to social threats with efforts to re-affiliate (Shilling & Brown, 2016). In some situations, social threats can produce behavioral changes that are combative or self-isolating, especially when paths to reaffiliation are unavailable or unachievable. Biologically, social organisms respond to social threats by upregulating inflammatory activity (Slavich & Irwin, 2014) in much the same way they respond to physical threats (MacDonald & Leary, 2005). One explanation for this phenomenon is that social threats, such as meeting a conflictual conspecific, greatly increase the impending risk for physical threats, such as physical wounding (Dhabhar, 1998; Slavich, 2020a, 2022). Historically, and especially without social support, injury could quickly turn into infection and death. Consequently, pro-inflammatory cytokine levels reliably increase for most people following laboratory-based social stress tasks (Steptoe et al., 2007), potentially functioning to prepare the body for injury or infection.

Highlighting the highly conserved nature of this dynamic, this inflammatory response to social stress is present in both humans (Slavich, Way, Eisenberger, & Taylor, 2010; Slavich & Irwin, 2014) and non-human animals (Sapolsky, 2005). For example, in one animal model study, mice exposed to social stress (i.e., social isolation followed by regrouping), but not other types of stress (e.g., physical restraint or loud noises), exhibited greater pro-inflammatory cytokine reactivity to an experimental inflammatory challenge (i.e., lipopolysaccharide administration; Gibb et al., 2008), highlighting the unique relevance of social threats for upregulating inflammatory activity in social animals. As discussed in greater detail by Slavich (2020a), intermittent increases in systemic inflammatory activity to social threat are adaptive and may not pose biological harm. However, as humans have evolved the ability to recall past social experiences, imagine potential future interactions, and take others’ perspectives to make inferences about the perceptions of others, these imagined social interactions, when viewed in a negative or threatening light, can initiate the same immunological responses as the real ones, making humans at risk for experiencing psychologically mediated, persistent inflammation (Slavich & Cole, 2013), a type of so-called sterile inflammation (i.e., inflammation activated by non-microbial signals; Chen & Nuñez, 2010). Although, today, social conflicts are generally less likely to be resolved by physical combat and isolation from one’s social group is less likely to pose an immediate threat to survival, this evolutionarily conserved mechanism persists. Given the health implications associated with elevated inflammation in modern contexts (Furman et al., 2019), this inflammatory response to social threat could be considered an evolutionary mismatch (Williams & Nesse, 1991) given that our environments have changed dramatically in the last few hundred years, but evolutionary change is slow. Consequently, our biopsychosocial responses to social threats are not optimized for or adapted to our current social-environmental circumstances but, rather, the circumstances our ancestors faced hundreds and even thousands of years ago.

Social Safety Theory provides a biologically plausible explanation for how these highly conserved evolutionary dynamics can be helpful in the short term but increase the risk for chronic disease and premature aging when the neural and immune threat responses are persistently activated. Over
time, for example, frequent or chronic experiences of social threat— even if only imagined— can modify biological processes, such as glucocorticoid-driven inhibitory processes (Slavich et al., 2010). Although glucocorticoids typically downregulate inflammatory activity following stress, chronic stressor exposure is associated with decreased glucocorticoid receptor sensitivity that promotes chronic systemic inflammation (Avitsur et al., 2006; Cohen et al., 2012; Miller et al., 2002; Miller et al., 2014). Chronically elevated inflammatory activity, in turn, increases the impact of experiencing social threats in both humans (Miller et al., 2009b) and non-human animals (e.g., rhesus macaques; Cole et al., 2009), and also increases the risk of biological aging and early mortality (Furman et al., 2019). As described by Slavich (2020a), exposure to social threats can also lead to chronic inflammation by increasing social stress generation (i.e., navigating the world in a way that engenders social stressors) and modifying social safety schemas (i.e., believing that one is not socially safe, or that others are a source of threat as opposed to safety). In turn, these cognitive and behavioral responses can cause epigenetic modification of the glucocorticoid receptor gene in the neural transcriptome, inducing epigenetic reprogramming of innate immune cells and altering the hematopoietic output of these cells from the bone marrow, and/or triggering increased arborization of SNS fibers in the lymph node, which expands the neural–immune regulatory pipeline and can promote sustained neuro-inflammatory sensitization to social adversity.

As research on these mechanisms is relatively scarce, future studies are needed to investigate how humans respond to different types of social threats— behaviorally, psychologically, and especially biologically— to document how these responses differ across various types of threats. From the perspective of Social Safety Theory, we would hypothesize that social threats will be more salient and impactful for individuals with more negative social safety schemas as compared to those with more positive social safety schemas and, similarly, for those with less perceived social safety as compared to those with more perceived social safety. In general, we would expect to find that biological response patterns to all types of social threats will differ based on the specific characteristics of the threat and function of the resulting biological response, in addition to whether the response conferred an adaptive advantage throughout our evolutionary past. For example, most threats activate the SNS and HPA axis, which increases the availability of metabolic resources. This occurs because, over our evolutionary past, those who experienced an increase in the availability of metabolic resources in response to social or non-social threats were more likely to survive than those who did not mount this metabolic response. However, making metabolic resources rapidly available is likely maladaptive if the threat a person is facing is caloric deprivation. In this case, responding to caloric deprivation by quickly using up one’s stored metabolic resources would not have conferred an adaptive advantage for our ancestors. This is one example of how Social Safety Theory uses evolutionary logic to draw predictions about how people will respond psychologically, biologically, and behaviorally to different types of threats.

**Biology of Social Safety Theory from a developmental perspective**

Social Safety Theory also provides a framework for understanding how social safety and threat affect biological functioning and health on the scale of a person’s lifetime, especially as a function of experiences occurring during specific developmental periods. As described above, the HPA axis likely plays a key mechanistic role linking experiences of social safety and threat with health, given the involvement of the HPA axis in helping prepare the body for potential harm in the face of threat (Townsend et al., 2011). One way the HPA axis does this is by stimulating the release of the hormone cortisol, which is commonly known to activate a ‘fight or flight’ response to threat. Past research suggests that differences in HPA axis activity supported human adaption to our evolved environments (Spencer, 2017). Therefore, HPA axis and other biological system dynamics that effectively mitigated harm were highly conserved (Del Giudice et al., 2011; Slavich et al., 2010a).

Stress responsivity dynamics appear to be especially malleable during a few key developmental periods— namely, prenatal, childhood, and puberty. Canalization of HPA axis reactivity due to
stressor exposure during these periods can significantly affect individuals’ biopsychosocial responses to social threat in later life. Indeed, an analysis of eleven human studies found that all reported an association between prenatal stressor exposure and HPA function in the child (Glover et al., 2010; see also O’Connor, Bergman, Sarkar, & Glover, 2013), with similar effects being found in animal model research (Creutzberg et al., 2021). This prenatal programming is thought to improve adaptation to the environmental context a child is likely to experience post-birth (Viltart & Vanbesien-Mailliot, 2007). Consequently, several affective conditions that are indicative of having experienced social threat have been found to be associated with aberrant HPA axis activity, including anxiety (O’Connor et al., 2005), depression (Diego et al., 2004), and posttraumatic stress disorder (PTSD) (Yehuda et al., 2005; see Tollenaar et al., 2011).

**Early-life Calibration of the Biological Social Stress Response**

Childhood is a particularly sensitive period of development during which time increased attention paid to social cues is critical for optimum development and survival. It is therefore an important period of investigation for understanding stress biology and testing Social Safety Theory. Early-life environmental conditions characterized by social safety and threat have been hypothesized to create a blueprint and set of expectations for the environmental conditions an individual is likely to encounter across their life course (Belsky & Pluess, 2013; Ellis et al., 2012; Frankenhuis & Del Giudice, 2012). Consistent with this possibility, threatening experiences occurring in childhood have been associated with the long-term dysregulation of multiple psychobiological systems independent of subsequent stressor exposure (Danese et al., 2009; Heim et al., 2019; Miller et al., 2011). Although these dynamics can damage health in the long term, they are designed to calibrate each individual’s biology to the unique demands of the person’s environment to increase his or her odds of reproduction and survival (Bush & Boyce, 2014). Consistent with this model, growing up in social-environmental conditions fraught with threat has been shown to predict the development of a ‘defensive’ phenotype (Miller et al., 2009a; Zhang et al., 2006) characterized by changes in neurocognitive (Hoffmann et al. 2018), HPA axis (Lé-Scherban et al., 2018; Roubinov et al., 2018; Zhu et al., 2019), hypothalamic-pituitary-gonadal (HPG) axis (Deardorff et al., 2014; McDade et al., 2016; Sun et al., 2017), and immune system functioning (McDade et al., 2016; Miller & Chen, 2010; Miller et al., 2011). These adjustments are theorized to enable an organism to efficiently use the resources available to defend itself from potential future threats, including violent others and pathogens. Consistent with this formulation, a recent meta-analysis of 29 studies and 4,292 participants, ranging in age from 8–62 years old, found that early life stress was related to blunted cortisol reactivity, and that this effect was stronger for adults than for children and adolescents (Bunea et al., 2017). Likewise, research has shown that adults exposed to moderate abuse or neglect during childhood develop particularly sensitive HPA axes, as indexed by potentiated cortisol responses to stressors (Raymond et al., 2021; Vegt et al., 2009). Moreover, the HPA axis reactivity pattern that emerges appears to be strongly shaped by the severity of childhood maltreatment experienced (Del Giudice et al., 2011; Vegt et al., 2009), with studies showing that individuals exposed to more chronic or severe abuse and neglect as children exhibit blunted diurnal cortisol rhythms (Vegt et al., 2009) as well as blunted cortisol reactivity to threat as adults (Carpenter et al., 2009; Raymond et al., 2021).

There is also some evidence that childhood threat exposure shapes biological stress reactivity during childhood. For example, children exposed to harsh parenting have been shown to exhibit greater cortisol responsivity to a social stressor (Bugental et al., 2003), whereas youth exposed to institutionalization, foster care (Koss et al., 2016), and abuse (Peckins et al., 2015) show blunted cortisol reactivity. Extending this work, Shakiba et al. (2020) found that children who experienced minimal or severe childhood adversity as indexed by restrictive parenting, family stress, and family socioeconomic status (SES) exhibited greater sympathetic and adrenocortical reactivity than those who experienced moderate childhood adversity. However, they also found that children who grew up in the most severe conditions exhibited blunted adrenocortical reactivity.
In addition, past research has shown that individuals who lack social safety in childhood, as characterized by exposure to early life abuse, neglect, or maltreatment, exhibit heightened inflammatory reactivity to social stressors, which, as alluded to above, is adaptive in response to actual threat but can have health-damaging effects over the long run (Carpenter et al., 2010; Schreier et al., 2020). So far, much of the research on childhood adversity and inflammation has focused on childhood SES (e.g., Miller et al., 2009a). This association makes sense according to Social Safety Theory, but additional research is needed to investigate whether such inflammatory responses are more strongly predicted by early life stressors that would have more greatly increased the risk of wounding over the course of evolution, such as childhood abuse, maltreatment, and neglect.

Applying a Social Safety Theory lens to stress biology may also help researchers interpret mixed findings that have pervaded the childhood adversity and biological stress reactivity literature. For example, there is quite a bit of research examining the impact of childhood sexual abuse on stress reactivity, but the results are mixed, with some studies documenting a significant impact of sexual abuse on stress reactivity and others not (Wesarg et al., 2020). One possibility is that individuals who have experienced early life abuse but do not exhibit aberrant biological stress reactivity profiles possess ample social safety in their lives (e.g., high maternal warmth or high-quality social support), do not have negative social safety schemas that drive downstream threat responding, or both. Viewing the childhood stressors themselves through the Social Safety Theory lens also may help make sense of these biological data. For example, perhaps abuse experiences that are most impactful for stress biology are those wherein the perpetrator was a close other or the abuse was ongoing, both of which we would hypothesize should lead to a more pro-inflammatory phenotype. Ultimately, the categories of childhood maltreatment, abuse, and neglect encompass a wide variety of actual experiences, some of which may be more relevant for the social signal transduction pathways that underlie Social Safety Theory than others. Re-categorizing childhood stressors through the lens of Social Safety Theory may thus help address these mixed findings and pinpoint the stressor characteristics that are most deleterious for health (Hamlat et al., 2022; Raio et al., 2022).

**Stress Responses During Puberty and Beyond**

In addition to the vast literature indicating an increased malleability of stress response pathways during childhood, there is burgeoning evidence that puberty is also a highly significant developmental period for restructuring of the biological stress response (Piekarski et al., 2017; Romeo, 2010; Slavich & Sacher, 2019). This is thought to be due to the evolutionary importance of puberty in facilitating successful survival and reproduction (Schooling, 2015). Along parallel lines, neuroscientists have documented structural and functional changes in neural circuits underlying the ‘social brain’ during this developmental period that make adolescents hypersensitive to social cues and experiences (Blakemore & Mills, 2014).

**HPA axis**

Recent research on this topic has found evidence that the effects of childhood adversity on cortisol responsivity and health are moderated by pubertal status (King et al., 2017), indicating a possible role for sex hormones in modulating the early life adversity-HPA axis link. Along these lines, several studies have found that early life stress, pubertal timing, and psychiatric diagnoses including depression and PTSD interact in predicting both diurnal cortisol patterns and cortisol responses to stressors (e.g., King et al., 2017; Negriff et al., 2021). Consistent with Social Safety Theory, some theorists have proposed that social threats may be especially relevant in this regard (e.g., Ellis et al., 2022). Given that limited research has directly compared the effects of different types of early life stressors, however, how social threats occurring around the pubertal transition restructure HPA axis dynamics that have implications for health remains unclear.
**Oxytocin system**

HPA axis dynamics are critical for understanding how social threats occurring during the pubertal transition and beyond affect lifelong health and behavior, but the HPA axis is not the only relevant system. Indeed, the oxytocin system also plays a key role in regulating social stress-related processes. Oxytocin is a neuropeptide hormone produced by magnocellular neurons of the paraventricular nucleus in the hypothalamus. Although best known for its role in shaping parenting behaviors such as lactation and labor, oxytocin is also thought to mediate the health benefits of warm social ties (Porges, 1998, 2003; Uvnäs-Moberg, 1998). Oxytocin has historically been overlooked in research on the developmental calibration of stress response systems. However, there is now increasing evidence that the oxytocin system is directly involved in evolved survival (e.g., resource accumulation, reproduction), and that its function is highly malleable during critical developmental transitions, such as puberty (Quintana & Guastella, 2020).

Quintana and Guastella (2020) reviewed most of the existing research on oxytocin and concluded that there is evidence that the oxytocin system functions similarly to any other evolved system, restructuring itself during different developmental periods and when faced with major biological events such as puberty. Specifically, it is hypothesized that oxytocin aids in the formation of social affiliation, which can, in turn, down-regulate stress reactivity (Heinrichs et al., 2003; Taylor et al., 2000). Although the covarying association between oxytocin and cortisol is understudied, there is some evidence that greater oxytocin reactivity to laboratory-based stressors predicts post-stressor declines in cortisol (Alley et al., 2019). Moreover, higher oxytocin levels in young adults have been associated with better accuracy in processing cognitive-emotional information during an acute stressor and more positive effect during exposure to a chronic stressor (Young Kuchenbecker et al., 2021), suggesting a protective role for oxytocin during times of stress following puberty.

When considered in the context of Social Safety Theory, one possibility is that when faced with a social threat, individuals exhibiting higher oxytocin levels may experience greater social safety – or, conversely, that experiencing greater social safety may lead to increases in oxytocin – both of which could mitigate the negative effects of social threat on health and well-being. This beneficial effect could occur through the cognitively mediated pathways described here (e.g., more health-promoting thoughts during times of threat) but there is also growing evidence indicating that oxytocin affects inflammatory activity, meaning that the oxytocin system may potentially mitigate the negative impact of social threat on health by modulating the immune system. Consistent with this possibility, both animal model and human research has found that oxytocin affects inflammatory dynamics via its interactions with microglia (Karelina et al., 2011; Yuan et al., 2016) and macrophages (Li et al., 2016; Szeto et al., 2017). Research has yet to investigate how associations between oxytocin and the immune system change across development and, in turn, affect risk and resilience to social threat. Given the synergistic goals of the oxytocin and immune system, though, pursuing this research is a natural next step. Specifically, research is needed to characterize dynamic associations between oxytocin, cortisol, and inflammation to determine how these systems interact across different social-environmental experiences, such as during exposure to social threat without the possibility of social safety, exposure to social threat with the possibility of social safety, and exposure to social safety cues or safe people without the presence of social threat.

**Immune system**

Finally, the immune system represents yet another key social signal transduction pathway that can be influenced by a variety of developmentally determined events and contexts (Slavich & Auerbach, 2018; Slavich & Irwin, 2014). These effects begin as early as perinatal development, during which time infections and elevated systemic inflammation may sensitize the developing brain and immune system to later injury and chronically elevated central nervous system (CNS) inflammation (Hagberg et al., 2012). Critically, there is also evidence that elevated perinatal inflammation may contribute to poor social functioning and neurodevelopmental delays in childhood (Giraud et al., 2020;
Girchenko et al., 2020) in what could become a developmentally sensitive positive feedback loop that affects immune system function, health, and development over time. Further, infection and inflammation are common causes of preterm birth (Goldenberg et al., 2008), which itself is associated with negative social sequelae that may shape the development of social safety schemas (Dean et al., 2021; Ritchie et al., 2015).

The bidirectional effects of inflammation on social safety and threat are likely present at all stages of development, but there is evidence that these regulatory effects may be especially strong following the pubertal transition. Indeed, inflammatory dynamics change across adolescence (Mac Giollabhui et al., 2021), and pubertal status is associated with some inflammatory cytokines [i.e., tumor necrosis factor-α (TNF-α) and interleukin-8 (IL-8)] above and beyond age alone (Stumper et al., 2020). These results indicate that puberty, a known period of social reorganization and reprioritization (Brown & Larson, 2009), may directly affect inflammatory biology. Given that inflammatory activity is associated with enhanced sensitivity to both negative and positive social feedback (Muscatell et al., 2016; Slavich et al., 2010), and that social evaluation, devaluation, and rejection trigger increased inflammatory activity (e.g., Dickerson et al., 2009; Giletta et al., 2018; Murphy et al., 2013), it is plausible that inflammation and experiences of social safety and threat may dynamically co-regulate each other, especially following the pubertal transition when sex hormone-enhancement of pro-inflammatory cytokine activity is stronger (Slavich & Sacher, 2019). If confirmed, this dynamic might explain some well-established gender differences in negative health outcomes that have been associated with interpersonal stressor exposure and inflammation, such as depression (Hankin et al., 1998; Slavich et al., 2020). Moreover, one might expect that this bi-directional association is particularly strong in the context of highly salient adolescent peer relationships (Brown & Larson, 2009).

The co-regulation between immune-inflammatory development and social processes that strengthens following puberty likely induces effects that persist over the life course, but which also get occasionally remodeled as a result of significant life events. At the same time that normative age-related worsening of immune function (i.e., immunosenescence) occurs, for example, so too does the likelihood of experiencing major changes to social relationships and dynamics that can affect individuals’ social safety schemas. On the one hand, older age can bring with it the death of loved ones and age discrimination, which promote experiences of social isolation. On the other hand, the birth of grandchildren, increasing support from younger generations, and greater recognition and respect at work and in the community can foster experiences of greater social status, inclusion, and belonging. In this way, social safety schemas may well moderate the process of immunosenescence and, consequently, disease burden and mortality risk as individuals transition into older adulthood.

Looking forward, additional research is needed to investigate the reciprocal regulation of social signal transduction pathways and experiences of social safety and threat across development, ideally using a mix of study designs across the life course. Quasi-experimental designs planned around common social transitions (e.g., transition to college) or periods of accelerated biological change (e.g., puberty) will provide useful information about naturalistic, developmentally anchored covariation. Furthermore, given evidence of bidirectional effects between biology and behavior (Dickerson et al., 2009; Lopes, 2017; Moriarity et al., 2020; Muscatell et al., 2016), tests of bidirectionality should be prioritized. When possible, it will be important to test within-person effects that are suggestive of causal relations (Falkenström et al., 2017) in both observational and intervention research to maximize the quality of the resulting data and the clinical relevance of this work. Additionally, it will be imperative to consider temporal specificity (i.e., the ideal time lag between measurements to capture an effect) when planning longitudinal studies on this topic (Moriarity & Alloy, 2021). Particularly when developmentally sensitive effects are possible, months – or years-long follow-up periods, which are common in psychological research, increase the risk of confounding, meaning that shorter follow-up timeframes are preferred. Finally, as in other areas of research, studies examining the biology of Social Safety Theory should use experimental designs whenever possible to more precisely investigate how social-environmental, neurocognitive, and biological processes interact to shape health, behavioral, and aging outcomes, and how these interactions might change over different developmental periods.
Social-cognitive aspects of Social Safety Theory from a developmental perspective

Social Safety Theory also provide a framework for understanding how social-cognitive processes, including social safety schemas, develop and change over the life course, and, in turn, affect human biology, health, and behavior over time.

Development of Social Safety Schemas

Consistent with research on the development of other cognitive schemas (Kaslow et al., 2000), we posit that social safety schemas are structured and restructured most prominently during developmentally and biologically sensitive time periods, such as infancy, childhood, the pubertal transition. We hypothesize that major life stressors can also greatly impact social safety schemas, especially when they critically change social perceptions, structures, or bonds. In each instance, social events can induce changes in social-cognitive beliefs but so, too, can the meaning and messages that individuals encounter, especially from parents, caregivers, and close friends (Slavich, 2020a, 2022). For example, different parenting styles can either encourage or discourage a child's belief in their ability to successfully navigate social situations (e.g., 'You can make friends' vs. 'You are a loner'), fostering self-schemas characterized by social safety vs. threat, respectively. Similarly, early life experiences can set the stage for social safety schemas related to the social world (e.g., 'Your friends will support you' vs. 'Other people are not trustworthy') and future (e.g., 'You will have a lot of friends who care about you' vs. 'You will be alone'). In each instance, these schemas result from both direct interactions, such as listening to a caregiver's positive vs. negative commentary, and from indirect observation, such as watching parents get divorced or acting coldly or hostilely toward each other.

As elaborated upon below, the formation and subsequent updating of social safety schemas are also likely influenced by broader socioenvironmental factors, including the relative safety of a person's neighborhood (Simons & Burt, 2011). Additionally, a key factor influencing the development of schemas oriented toward the expectation of social threat are adverse childhood experiences (ACEs) such as early-life abuse and neglect, which, beyond having the above-described biological effects, can lead to the internalization that others are unpredictable, undependable, untrustworthy, and dangerous (Neil et al., 2022). In addition to direct and indirect experiences, childhood attachment styles and their interplay with the social environment could shape an individual's social safety schemas. Whether attachment styles and social safety schemas are distinct constructs, overlapping, or identical, however, requires further investigation.

Whereas childhood may shape the initial development of social safety schemas, adolescence is likely a period during which time these schemas are substantially developed and reorganized. Not only is adolescence a period of increased salience of peer relationships and social autonomy (Brown & Larson, 2009), it is also when other schemas stabilize in terms of both factor structure and temporal stability (LaGrange et al., 2008). First, in adolescence, the increased salience of social relationships and freedom to interact with peers without supervision provides opportunities for more frequent social experiences that can impact social schemas (Brown & Larson, 2009). Second, in addition to relationships with others, adolescence – particularly late adolescence – is a period marked by heightened goal-directed activity (e.g., application for first jobs or college) that could foster schemas related to social safety and threat, both directly through acceptance or rejection, as well as indirectly through support, or lack thereof, around these key life events. Third, the transition to adolescence is characterized by a shift from self-identifying through concrete, behavioral characteristics to more abstract concepts (Harter, 1986, 1990). Fourth, this developmental transition is also characterized by increased consideration of past experiences to inform interpretations of the present and expectations for the future (Rholes et al., 1980). Fifth, mid-childhood and adolescence is a critical period of risk for the development of internalizing disorders, which operate in positive feedback loops with negative schemas (Alba & Calvete, 2019; Gómez-Odriozola & Calvete, 2020). Sixth, as
late adolescents/emerging adults transition away from their parents and into new living environments such as dorm rooms and apartments, there is a plethora of opportunities to internalize beliefs about one's independent ability to successfully navigate social threat, as well as surrounding the nature of the social world and the likelihood of experiencing social safety vs. threat as an adult.

Looking forward, additional research is needed to determine when social safety schemas solidify, when they are highly malleable and sensitive to social-environmental input, and which types of stressors and life experiences are most likely to have an impact. In terms of natural life events, there exist many developmental transitions across life course that require significant social change or adjustment, and we hypothesize that social safety schemas are likely most malleable during these significant transitions. In addition, there are many life events in adolescence and adulthood that can induce major social changes, including children gaining independence and leaving the house, divorce, deaths, and the transition into old age. Along these lines, the malleability of social safety schemas to naturally occurring life events (e.g., accepting and nurturing relationships, transition to college, peer rejection), laboratory-based stressors, and targeted treatment (e.g., cognitive behavior therapies) should be evaluated. Relatedly, studies are needed to investigate if a single social safety schema measure can be used for all individuals or if developmentally specific measures are necessary. In addition, research is needed to better understand how social safety schemas and stress biology co-create and bi-directionally influence each other. Finally, to maximize the clinical utility of Social Safety Theory, research should investigate if certain schemas (e.g., social safety vs. threat schemas; social self vs. world vs. future schemas) are more responsive to intervention than others. If evidence arises that a particular social schema is malleable, research could next evaluate it as a potentially modifiable mechanism linking social stressors with negative health-related processes and outcomes, including dysregulated metabolic and immune system function, anxiety, depression, and somatic/physical diseases.

The role of culture and social institutions in Social Safety Theory

In addition, Social Safety Theory provides a framework for investigating how complex social networks and connections, which exist within larger interacting sets of social circles and institutions, combine to affect stress biology and behavior. These institutions include families, schools, governments, and economic and legal systems (among others) that both constrain and encourage behavior through implicit and explicit guidance (Martin, 2004). Such guidance can come from norms or expectations, ideologies, and policies (Martin, 2004). As such, these institutions are deeply embedded in culture(s), or the internalized and shared framework through which both the individual and collective experience the world (Singer et al., 2016). Importantly, cultures and social institutions shape individuals and are, in turn, shaped by them (Singer et al., 2016), with this complex bidirectional dynamic having important implications for how social and biological processes underlying Social Safety Theory initially develop and are maintained over time.

According to Social Safety Theory, social institutions can affect the topography of social networks, and thus experiences of social safety and social threat, in several ways. For example, our involvement in schools and workplaces can affect who we encounter and subsequently befriend based on salient demographic categories such as race/ethnicity, sex/gender, and SES. Historic and ongoing exclusion of some individuals based on demographics from schools, workplaces, and other social institutions – whether explicit or implicit – can contribute to homophily in social networks (see McPherson et al., 2001). Homophily transcends biology, extending to occupations, education level, political and religious affiliations, and other categories (McPherson et al., 2001), and has implications for availability of social support. For instance, in her study of middle managers’ social networks in four Fortune 500 companies, Ibarra (1995) found that minority managers’ networks were characterized by lower homophily than their White counterparts. These managers also tended to have fewer close relationships in the workplace. Notably, the structural or institutional constraints on minorities, including underrepresentation and concerns about how minority homophilic associations would appear to
White colleagues, led to different social network structures among minority managers to achieve the same goal – namely, career advancement (Ibarra, 1995) – all of which has implications for experiences of social safety and threat over time.

Structural constraints driven by interactions with social institutions can also affect the type and amount of social support and threat present in social networks. Prior to Hurricane Katrina in New Orleans, for example, disadvantaged residents would have needed to know 80 more neighbors than their wealthier counterparts to achieve the same likelihood of pre-storm social assistance (Elliott et al., 2010). During the storm and in its immediate aftermath, as individuals shifted to trans-local social connections, low-income residents were at similar disadvantage due to fewer trans-local ties, an inability to engage with those ties due to financial constraints, and the likelihood that the ties were similarly disadvantaged (i.e., economic homophily; Elliott et al., 2010). Additionally, geographic areas with greater social capital can mobilize to achieve a greater share of disaster relief money and otherwise improve their chances of recovery (Aldrich, 2017). In the absence of significant disaster relief, low-income residents may need to continue to rely on their social networks for support and recovery. This could place significant strain on these social ties, particularly in times of repeated crises, even absent major events such as hurricanes or pandemics. What should be evident from these examples is that factors that sustain social safety schemas – and, indeed, individuals’ social realities – are not confined to the individual (i.e., their personality, genetics or life history) but rather extend beyond the person to include macro factors such as neighborhoods, institutions, and policies that exert a significant impact on lifelong health and opportunity.

The durability and persistence of social institutions derives from their ability to create stable expectations about the behaviors of others (Hodgson, 2006). In this regard, institutions may themselves contribute to social safety to the extent that such reliable expectations are welcome. On the other hand, expected behavior could be negative, as is the case of systemic racism and discrimination, and contribute to actual or perceived social threat. Such higher-level sources of collective social threat likely interact with acute and chronic social stressors that occur directly to people or within their social network(s). For instance, the effects of prolonged racism or discrimination could be exacerbated by belonging to a network characterized by low social safety and high social threat; conversely, these effects could be mitigated by a social network that has high social safety and low social threat. Consistent with this possibility, daily discrimination has been found to be related to higher C-reactive protein (CRP) levels and fibrinogen in Black adults (Surachman et al., 2021), whereas discrimination and other forms of social threat have been related to shorter telomere length, a biological marker of aging (Chae et al., 2014; Damjanovic et al., 2007; Epel et al., 2004; Mayer et al., 2019; Wilson et al., 2019). Considering higher-level sources of social safety and threat and the role they play in shaping social networks and experiences – both alone and in concert with individual-level risk and resilience processes – is an important avenue for testing and extending Social Safety Theory.

The role social institutions play in creating stable behavioral expectations may create another source of higher-level social threats and stressors – namely, instances when behaviors do not accord with expectations. This could be considered analogous to predictive processing theories of cognition, wherein the brain actively generates and updates predictions about sensory input based on prior inputs and predictions in a Bayesian fashion (Friston, 2010; Friston et al., 2014; Kube, Schwarting, Rozenkrantz, Glombiewski, & Rief, 2020). Prediction errors, or mismatches between predictions and sensory inputs, are used to update expectations (i.e., shift priors to posterior probabilities; Greenhouse-Tucknott et al., 2022); however, they may also underlie risk for stress-related outcomes, including pathological fatigue and depression (Stephan et al., 2016) as well as allostatic load (Barrett & Simmons, 2015).

Similarly, prediction errors related to expectations derived from social institutions (e.g., during the COVID-19 pandemic) may contribute to feelings of social threat and concomitant allostatic load as an individual attempts to reorient themselves in a social environment that changes faster than social institutions can cope (Razavi et al., 2020; Slavich et al., 2022). The process of acculturation is
another potential source of mismatch between internalized models of expected behaviors and actual actions taken by others (Gendron et al., 2020). Applying Social Safety Theory to individuals or contexts undergoing rapid social change may help to greatly expand our understanding of how health, well-being, aging, and behavior are influenced by macro-level factors that promote social safety and threat, whether these social changes are caused by disruptions to social networks, changes to social institutions or policies, or both. For example, globalization, conflict, disaster, and climate change are each actively causing significant human migration and distress, fraying well-established and protective social networks and institutions (Segal, 2019). The consequences of such large-scale social change are likely greater for minoritized groups, with a greater risk for negative health outcomes. Finally, taking a bird’s-eye view of social safety and threat and considering how individuals are situated within social structures and institutions allows us to integrate Social Safety Theory with wider research on health disparities. We expand on these ideas below to describe in greater detail how Social Safety Theory can help explain health disparities in vulnerable populations.

**Explaining health disparities in stigmatized, marginalized, and disadvantaged populations through the lens of Social Safety Theory**

Social safety and social threat are particularly relevant for understanding the health and well-being of oppressed and marginalized populations who often experience discrimination, aggression, rejection, and exclusion. Early scholars described oppression as dehumanization, which influences isolation and hinders access to education, employment, and social status of less well-represented populations (Freier & Kletter, 1970). From the perspective of Social Safety Theory, the presence of oppression, marginalization, and associated experiences of discrimination (i.e., the ‘unjust or prejudicial treatment of a person or group of people’) are forms of social threat that clearly entail a lack of social safety (Lewis & Van Dyke, 2018, p. 176). These experiences are thus hypothesized to activate the Biobehavioral Response to Social Threat, which mitigates physical and microbial risk following injury but has negative long-term consequences for health.

Although few studies have explicitly tested Social Safety Theory in diverse populations, many studies have investigated how discrimination and other oppressing experiences negatively impact health in marginalized communities. Broadly speaking, this research has shown that experiences of discrimination (e.g., racism, sexism, homophobia, transphobia) can significantly affect the health, health behaviors, and mortality risk of marginalized populations. We summarize this work below by examining how social threat affects two key marginalized groups: racially and ethnically diverse populations, and sexual and gender diverse populations. We then conclude the section with a discussion of intersectionality and the importance of measuring intersecting identities in future research. These processes are not specific to the below-discussed groups but we highlight them as they are well-studied effects and a prototypic example of how experiencing social adversity across the life course degrades health and well-being.

**Racism and Health**

One of the most predominant ways in which oppression manifests is racism. Racism is deeply ingrained in the United States and is one of the most prevalent forms of oppression (Carter et al., 2017). Although studies applying Social Safety Theory to racially and ethnically diverse populations are lacking, the theory provides an explanatory framework for understanding existing findings in this literature. For instance, extensive research has examined how racial and ethnic marginalization (e.g., discrimination, aggression, rejection, exclusion) predicts poor health in a variety of marginalized populations. In reviewing this literature, Carter and colleagues (2017) conducted a meta-analysis of 105 studies on racial discrimination and health and found that racial discrimination was consistently related to poorer mental and physical health. Research also has found that discrimination
increases the risk of experiencing several major health problems, including hypertension, breast cancer, obesity, and high blood pressure (see Williams & Mohammed, 2009).

Similarly, Pascoe and Smart Richman (2009) conducted a meta-analysis of 110 studies investigating the association between discrimination and mental health. Of the 500 effects examined, 90% indicated that discrimination is related to experiencing more mental health problems and 69% of these associations were statistically significant. Some researchers have hypothesized that these mental health disparities arise from greater substance use in racially and ethnically diverse populations, but substance use may also be due to discrimination exposure insofar as substance use is one way that individuals cope with the harm of marginalization (Bennett et al., 2005). In short, discrimination likely impacts the mental and behavioral health of marginalized individuals through multiple biobehavioral pathways that are delineated by Social Safety Theory.

Discrimination also negatively impacts the physical health of people of color—an effect that is exacerbated by the current structure of the U.S. healthcare system, which promotes collective experiences of social threat and a sense that there is a general lack of safety. Along these lines, extensive research has shown that communities of color have poorer access to quality and reliable health care, and they are often dissatisfied with their care (Riley, 2012). This lack of access to quality care can contribute to greater mortality in populations of color and was an especially significant problem during the COVID-19 pandemic. For example, in the United States, mortality due to COVID-19 was up to three times higher in Black and Latinx communities than in White communities (Alcendor, 2020).

Discrimination and Health

Although racism is a major problem in the United States that contributes to significant health disparities in populations of color (Trent et al., 2019), many other social threats also impact health, including gender and/or sexual orientation-based discrimination. In fact, sexually diverse populations, such as those who have had sexual or romantic relationships or attractions outside of exclusive heterosexuality, and gender diverse populations, such as those with gender identities or expressions that differ from their birth-assigned sex/gender, who may identify as trans, non-binary, and gender fluid (SGD), are known to experience more health disparities than their heterosexual counterparts (Diamond et al., 2021). Substantial research has documented mental health disparities in SGD populations, with heterosexual individuals having far better mental health than sexually diverse and bisexual individuals, with fluid individuals faring the worst (Frisell et al., 2010; Hatzenbuehler, 2014). Consistent with Social Safety Theory, researchers have hypothesized that these disparities are likely a result of cumulative minority stress, or the experience of prejudice, rejection, internalized homophobia, and hiding of one’s identity (Frisell et al., 2010; Hatzenbuehler, 2014; Meyer, 2003). Social Safety Theory advances this work by connecting the dots between experiences of social threat and health-damaging biological dynamics.

Studies have also shown that SGD populations face significant disparities in physical health. Consequently, many researchers have attempted to extend Meyer’s (2003) minority stress theory to describe how both mental and physical health disparities develop. Specifically, Diamond et al. (2021) reviewed existing research on health disparities in SGD populations and used Social Safety Theory as a theoretical explanation for these disparities. In doing so, they suggested that minority stressors and discrimination signal a lack of social safety and the presence of social threat, which in turn activate biological responses meant to mitigate physical and microbial threats that can degrade health if prolonged.

Consistent with this model, past studies have found that sexual orientation/identity discrimination (e.g., homophobia) is a strong driver of physical health disparities in SGD populations (for a review, see Lick, Durso, & Johnson, 2013). In general, as compared with heterosexual individuals, SGD individuals exhibit elevated rates of arthritis, functional disability, asthma, some cancers, and cardiovascular problems (Diamond et al., 2021; Fredriksen-Goldsen et al., 2013). Many of these studies reflect the literature examining social threat and mental health, suggesting that bisexual
individuals – and primarily bisexual women – have the poorest health outcomes (Steele et al., 2009). Past research also has found that SGD populations have greater mortality risk than heterosexual populations, although this effect may be driven by sexually diverse women (Laughney & Eliason, 2022). Again, similar to the literature on racial and ethnic diversity, research suggests that elevated substance (e.g., alcohol and tobacco) use in SGD populations may be partly responsible for the greater health difficulties evident in SGD populations (Bränström & Pachankis, 2018; Cochran & Mays, 2009). Consistent with Social Safety Theory, however, this research again suggests that this increased substance use might not be due to SGD identities but rather to the high levels of discrimination experienced in this population (McCabe et al., 2010).

Although disparities in health exist between heterosexual and SGD populations, it is also important to note that significant disparities exist within the SGD population. For example, bisexual individuals experience more discrimination and have poorer health outcomes than their heterosexual and gay counterparts (Diamond et al., 2021; Hatzenbuehler, Phelan, & Link, 2013). Studies have suggested that bisexual individuals may exhibit the worst mental and physical health outcomes because they face social threat from not only heterosexual individuals but also other members of the SGD community (Meyer, 2003; Ross et al., 2018; Yoshino, 2000). Likewise, there is growing evidence that gender-diverse individuals face greater health disparities than their sexually diverse counterparts due to the unique discrimination experiences of trans and non-binary individuals in the United States (Diamond et al., 2021). Although access to safe and affirming medical care that fosters social safety has increased in recent years (Chen et al., 2018), this is still a major hindrance for gender-diverse individuals seeking care and presents difficulties when these individuals seek practitioners for gender-affirming surgeries and interventions. Ultimately, determining the variability in health outcomes attributable to social threat vs. poor access to high-quality care is an important topic that requires additional attention.

Intersectionality

Individual differences in health disparities within SGD populations bring up another critically important concept for future research: intersectionality. Intersectionality, or the influence of multiple intersecting identities (e.g., age, gender, sexual orientation, SES) on social interactions and environmental experiences, is critically important for understanding how complex, interrelated sources of social threat impact health. One of the most important health-relevant factors in this vein is SES, and Social Safety Theory provides a model for understanding how differences in SES impact health.

SES is considered one of the greatest drivers of health disparities across the world (WHO Health Commission, 2008). Whereas low SES itself can hinder an individual’s ability to engage in health-promoting behaviors and reduce access to quality food, medical care, quality education and more, it is challenging to separate SES from key health-damaging social threats, such as oppression, social inequality, rejection, and exclusion. One way to measure SES, for example, is by household wealth, and the racial differences in wealth in the U.S. are staggering. For every dollar earned by White households in the United States, Asian households earn 83 cents, Hispanic households earn 7 cents, and Black households earn 6 cents (US Census Bureau, 2014). These striking differences in income can lead to poorer health outcomes for these populations of color by multiple biopsychosocial mechanisms that include exposure to SES-related stigma, lack of resources for health care and health promotion, and lack of access to education. How intersectionality is related to these mechanisms, however, is complicated and dynamic. Take Black men, for example; they hold one marginalized identity (being Black) and one high-status identity (being male). Yet, Black men are disproportionately targeted by police, and this cannot be easily explained by simply examining the number of marginalized identities that this population holds (Hester et al., 2020). Therefore, it is important to investigate not just how many marginalized identities a person holds but how a
person’s marginalized identities interact with one another, and how these interactions, in turn, influence experiences of social safety and threat.

Intersectionality is also an important concept when examining health outcomes. For example, using cross-sectional data, Mereish and Bradford (2014) found a three-way interaction between race, gender, and sexual orientation in predicting substance use problems, suggesting that the intersection of these identities is nuanced and complicated, and that assessing all three factors may be important for understanding health disparities. More specific to Social Safety Theory, Van Dyke et al. (2017) found that greater SES-based discrimination predicted higher levels of CRP, but only for Black individuals who had higher educational status. In general, though, studies on intersectionality are seriously lacking. There is certainly value in understanding how specific social threats such as racial, ethnic, gender, and orientation-based discrimination predict health outcomes in specific marginalized populations, but that is not enough. Future research, we believe, should measure not only multiple identity facets (e.g., gender, race, ethnicity, disability status, sexual orientation) but also factors at both the micro and macro level that identify general trends in social attitudes. At the micro-level, researchers should consider measuring factors such as internalized racism and homophobia, family attitudes and acceptance, religious affiliation, personal SES, and social support. At the macro level, in contrast, researchers should consider measuring factors such as political affiliation in the geographic location in which people live, current political attitudes and trends, and policies that both increase and reduce oppression and exclusion in order to fully appreciate how the social environment shapes mental and physical health. Along these lines, Gonzales and Ehrenfeld (2018) found that SGD individuals living in states with fewer structural protections for SGD populations were more likely to report poor health. Again, however, work on this general topic is sorely lacking.

The main point here is that although substantial research has examined how social threats such as discrimination and exclusion affect health, very few studies have investigated how multiple intersecting identities lead to experiences of social threat that, in turn, influence disease risk, morbidity, or mortality. Furthermore, considerable research has examined associations between social threat and health outcomes in marginalized groups such as Black and SGD individuals as a whole, but very few studies have looked at these relations in detail to elucidate, for example, within-groups differences or population-specific risk and protective factors. In addition, many marginalized groups remain largely understudied, including multiracial individuals, persons with mental and physical disabilities, and individuals who go against prevailing social ‘norms’ such as women who choose to not have children and men who choose to be the primary caretaker. Further, many of the studies presented above used retrospective measures, and there is a distinct lack of longitudinal work aimed at elucidating the timing of these processes or providing evidence of causal associations. In each of these cases, applying Social Safety Theory provides not only a framework for predicting which specific types of adversity should be most impactful, but also a model for understanding how these stressors get converted into health-altering biology.

Moving beyond threat: The protective aspects of social safety

One important contribution Social Safety Theory makes, which we have perhaps only implicitly mentioned thus far, is that it provides a conceptual framework for understanding not just why socially threatening experiences are harmful to health but also why social safety appears to buffer the negative effects of social threats in health disparities research. In a recent prototypic study on this topic, researchers examined the association between discrimination, depression, physical health, and social support in 204 Latinx adult immigrants and found that the association between discrimination and physical health was mediated by depression; however, the presence of social support protected those who experienced discrimination from developing negative health outcomes (Cariello et al., 2022). Social Safety Theory can help explain these effects insofar as the presence of social safety would be hypothesized to downregulate SNS and HPA axis activity, which would, in turn, reduce...
inflammatory responding, a key mechanisms underlying both depression and physical health problems. Therefore, even in the presence of social threat (i.e., discrimination), social safety (i.e., social support) appears to play a critical role.

In addition to providing a mechanistic model of how positive social experiences influence biological processes that confer resilience, this focus on social safety can also refocus research efforts away from only studying disease biology and toward also investigating biopsychosocial processes underlying desirable outcomes, such as affiliation, affirmation, inclusion, belonging, meaning, and joy. Studying these processes would generate the much-needed evidentiary base for salutogenesis and also provide new avenues for understanding the biopsychosocial roots of thriving and resilience in marginalized groups, including SGD populations and beyond (Diamond & Alley, 2022). This work may also lead to the identification or development of additional strategies health care providers, policymakers, and others can use not only to reduce social threat and disease risk, but also to increase social safety, strength, and well-being in individuals and communities alike.

If we can ultimately identify social factors related to differences in health but also the biological processes through which these effects occur, we will come one step closer to developing more effective – and perhaps even personalized – interventions for reducing disease risk and promoting well-being. Such interventions are needed for all but especially for marginalized populations who experience the greatest burden and who need innovative, health-promoting strategies the most. Social Safety Theory provides a framework for conducting research on these topics insofar as it describes the evolutionary principles, dynamics, and pathways governing social regulation of the human brain and immune system, thus advancing our understanding of stress biology, and helping create a biology of psychosocial strength and wellness.

**Comparing Social Safety Theory with related perspectives**

As described in Table 2, Social Safety Theory can be compared with other perspectives on the conceptualization of stressors and stress biology. First, there are many similarities between Social Safety Theory and Attachment Theory (Bowlby, 1958), which is a conceptual predecessor to Social Safety Theory. A key difference is that Social Safety Theory extends beyond psychosocial processes affecting health to describe a complete set of biologically plausible mechanisms linking social experiences with human health and behavior. Second, Social Safety Theory shares some similarities with the Generalized Unsafety Theory (GUTS; Brosschot et al., 2017), but differs from GUTS with respect to its descriptions of threat preparedness as well as basal, stress-related biology. Whereas GUTS posits that basal biological states are designed for threat preparedness, for example, Social Safety Theory hypothesizes that threat-related biological processes should only be activated for individuals who have experienced their early-life environment as being socially threatening. Moreover, Social Safety Theory accords a central role for social safety schemas in mediating these effects, with negative health effects being most likely for people who possess negative social self, world, and future schemas. Further, Social Safety Theory argues that an individual’s basal biology is shaped by the brain and immune system’s drive to calibrate to the social environment, and describes the full set of biological mechanisms underlying their bidirectional communication.

Third, Social Safety Theory can be compared with Social Baseline Theory (SBT; Beckes & Coan, 2011), which, similar to Social Safety Theory, conceptualizes social bonds as a highly desired and beneficial aspect of human nature. One key difference is that whereas SBT accords central importance to the human brain’s assessment of interdependence, shared goals, and joint attention in social relationships, with social resources mitigating risk and minimizing the need for energy and effort output, Social Safety Theory focuses on the absence of social threat and presence of social safety as key signals that reduce the likelihood of physical injury and infection, thus minimizing the need for the Biobehavioral Response to Social Threat to anticipatorily mitigate pathogen and, therefore, survival risk. Fourth, is the Adaptive Calibration Model (ACM; Del Giudice et al., 2011), which is different from Social Safety Theory in two key respects: First, whereas ACM focuses
largely on the HPA axis as a key mediator of social-environmental effects, Social Safety Theory describes a variety of other biological pathways and levels of analysis linking social experiences with health and behavior. Additionally, whereas ACM focuses exclusively on threat, Social Safety Theory focuses on both safety and threat. Finally, as alluded to above, Social Safety Theory can be compared with the deprivation and threat model of early adversity proposed by Sheridan and McLaughlin (2014). In addition to the distinctions noted above, whereas the deprivation and threat model explicitly focuses on early adversity, Social Safety Theory is a life course model. In addition, whereas the deprivation and threat model is primarily grounded in neural data, Social Safety Theory accords a central role to bidirectional neural-immune communication in shaping social behavior and human health.

Table 2. Comparing Social Safety Theory and related frameworks.

<table>
<thead>
<tr>
<th>Comparison Framework</th>
<th>Key Similarities and Differences</th>
</tr>
</thead>
</table>
| Attachment Theory          | 1. Social Safety Theory builds upon decades of research on Attachment Theory to describe the impact that social safety and threat have on psychological and biological processes  
2. The theories are similar with respect to their focus on social schemas  
3. A key difference is that Social Safety Theory describes the full set of neural, physiological, molecular, and genomic processes linking the social environment and health  
4. In addition, attachment theory is more explicitly consistent with psychoanalytic theory than Social Safety Theory                                                                                                                                 |
| Generalized Unsafety Theory (GUTS) | 1. Social Safety Theory posits that basal biological activity is shaped by a drive for neural-immune calibration to the social environment and describes the specific mechanisms subserving neural-immune communication  
2. Whereas GUTS posits that the basal biological state is one of threat preparedness, Social Safety Theory hypothesizes that this is only true for individuals who experienced their early life environments as socially threatening  
3. Compared to Social Safety Theory, GUTS focuses more explicitly on unconscious aspects of generalized unsafety and threat preparedness                                                                                                                                 |
| Social Baseline Theory (SBT) | 1. Both Social Safety Theory and SBT conceptualize social proximity as a generally desired and beneficial aspect of human nature  
2. A main difference is SBT’s focus on economy of action, wherein a primary benefit of social relationships is to preserve metabolic and neural resources; in contrast, Social Safety Theory focuses on the reduction of social threat and risk of physical injury afforded by socially safe relationships as being a key driver of social connection  
3. Additionally, whereas Social Safety Theory focuses on social isolation as being a risk factor for solitary individuals, SBT describes this in the context of risk-dispersion among groups                                                                                                                                 |
| Adaptive Calibration Model (ACM) | 1. ACM focuses strictly on threat and unpredictability, both of which confer social threat in the Social Safety Theory framework; in addition, Social Safety Theory extends this model to include social safety as a distinct resilience factor  
2. Whereas the key pathway in the ACM is the HPA axis, Social Safety Theory describes several additional health-relevant systems and levels of analysis, including neurobiological, immunological, molecular, and genomic processes  
3. The ACM includes a specified taxonomy of four stress responsivity patterns resulting from different early life environments (i.e., sensitivity, buffered, vigilant, & unemotional) whereas Social Safety Theory focuses on the more general, early-life calibration of neural-immune responsivity to perceived social safety and threat  
4. ACM responsivity patterns include theorized sex-differences in the manifestations of individual patterns (e.g., agonistic vigilance for males and withdrawn vigilance for females) whereas Social Safety Theory does not                                                                                                                                 |
| Deprivation and Threat Model | 1. Compared to the deprivation and threat model, Social Safety Theory posits that experiences of social safety and social threat are separable, orthogonal dimensions of human experience  
2. Although both the deprivation & threat model and Social Safety Theory focus on threat, the conceptualization of threat in Social Safety Theory is primarily concerned with social experiences that historically increased the risk of physical injury or infection over the course of evolution  
3. Both frameworks are based in data from social and affective neuroscience, but Social Safety Theory ascribes central importance to bidirectional neural-immune communication in shaping social behavior and human health  
4. Whereas the deprivation and threat model focuses specifically on childhood adversity, Social Safety Theory adopts a life course perspective on human health and behavior
Factors affecting social safety and threat

The sections above describe the theoretical and definitional underpinnings of Social Safety Theory, as well as its underlying social and biological mechanisms and how the theory can be applied to investigate differences in health and well-being across the life course. Now, we turn our attention to describing nine factors that can moderate the activity of the social signal transduction pathways that link social safety and threat with health and behavior. These factors include situational factors – namely, childhood microbial environment, birth cohort, culture, and air pollution – as well as individual difference factors – namely, genetics and gene expression, sleep, diet, personality, and self-harm. This list is not exhaustive but, rather, is meant to highlight especially promising avenues of investigation based on existing theoretical insights and findings. We begin with the situational factors and then transition to the individual difference factors affecting social safety and threat.

Situational Factors

Childhood microbial environment

The childhood microbial environment may play an important role in shaping an individual’s experiences of social safety and threat, as well as their biopsychosocial response to social adversity. As described by Slavich (2022), the immune system functions like a dynamic learning system that uses Bayesian forecasting to prepare for the specific microbial threats to which a person is most likely to be exposed given their past and current surrounding microbial environment. This continual calibration of the immune system over the life course refines the functional dynamics, capacity, and regulation of each individual’s immunologic response to provide the greatest immunological and survival advantage possible (McDade, 2012; Rook et al., 2017). Without this complex calibration, immune system defenses would be poorly prepared to deal with the specific pathogenic realities of the person, thus hampering effective pathogen clearance and increasing the individual’s risk for widespread infection.

One factor that greatly affects the specific types and diversity of microbes to which a person is exposed is the childhood living environment. More specifically, rural environments contain numerous healthy microbes from animals, feces, and soil that help educate the immune system and strengthen its regulatory capacity. Today, however, 82% of people living in North America live in sanitized urban environments that include excessive antibiotics use, treated drinking water, and infrequent contact with farm animals, mud, and feces (United Nations, 2018). It may seem surprising that such environmental characteristics have much to do with social safety and social threat, but hygienic urban environments minimize the frequency and diversity of microbial exposures and thus provide fewer opportunities for the immune system to practice upregulating inflammatory activity when it is needed and downregulating inflammation when it is no longer required (Slavich, 2020a). In contrast, less hygienic early-life environments – which are associated with a greater diversity of microbial exposures during critical periods of immune system development – lead to more frequent acute inflammatory responses and opportunities to develop a tightly controlled, counter-regulatory anti-inflammatory response. As a result, growing up in a more sanitary environment can hamper the development of a tightly controlled inflammatory response, potentially giving social threats a greater likelihood of fostering systemic chronic inflammation that does not resolve over time.

Supporting this possibility, McDade et al. (2013) recruited 1,622 young adults in the Philippines and studied the extent to which the effects of social adversity on systemic inflammatory activity were moderated by childhood environment. They found that parental absence resulting from divorce, separation, or death was associated with 17% lower CRP levels for individuals exposed to high levels of feces during infancy. In contrast, those exposed to low levels of animal feces during infancy exhibited 47% higher CRP levels in young adulthood relative to their no-stress counterparts. These data thus
indicate that social threat-related inflammatory levels are strongly moderated by early childhood microbial environment. In a more recent experimental study, Böbel et al. (2018) found that participants who grew up in a rural environment surrounded by animals exhibited greater HPA axis, SNS, and cardiovascular responses to an acute laboratory-based social stressor. These participants also found the stressor to be more anxiety-provoking, challenging, and threatening. When it came to participants’ inflammatory responses to the social stressor, though, it was those who grew up in an urban environment who exhibited the greatest social stress-induced increases in the pro-inflammatory cytokine interleukin-6 (IL-6), suggesting that counterregulatory processes that help limit inflammatory responses to social threat are stronger in individuals who grow up in microbiologically rich environments.

Data such as these are intriguing and shed new light on the complex interplay between social, environmental, and biological factors that are implicated in human health and disease risk. However, very few studies have combined assessments of childhood microbial environment with measures of lifetime stressor exposure and basal inflammatory levels or social stress-induced inflammatory reactivity profiles. Looking forward, therefore, it will be important to compare participants who have grown up in rural vs. urban environments to better understand how their antiviral and pro-inflammatory responses vary across different social threats, and to examine the immunoregulatory mechanisms underlying these differences that predict risk and resilience for disease.

**Birth cohort**

Like childhood microbial environment, the relevance of birth cohort for Social Safety Theory may not be immediately apparent. However, birth cohort is a strong predictor of not just social norms, experiences, and institutions, but also the immunologic and medical environments in which people are raised. According to the Organization for Economic Co-operation and Development (OECD, 2011), for example, the number of children born outside of marriage and without a solid family structure tripled from 11% in 1980 to nearly 33% in 2007. Likewise, divorce rates doubled and are now at 2.4 divorces/1,000 people, with approximately 10% of children living in reconstituted families and 15% living in sole-parent families. The likelihood of being exposed to explicit social threats has also risen over this time. In a recent UNICEF (2014) survey, approximately 70% of children 2–14 years old reported having experienced direct social threats in the form of verbal aggression, intimidation, humiliation, or ridicule, and approximately 60% reported having experienced physical threats such as hitting, kicking, spanking, or shaking.

Whereas socially protective norms have generally degraded over time, other social-environmental characteristics have improved, which bodes well for those born in the (later) twentieth century and 21st century. These individuals have benefited from better public health and hygiene, living standards, vaccine availability, and medical and dental care. Considered together, these improvements have reduced the likelihood of being exposed to inflammation-promoting pathogens; becoming infected in early life with bacteria such as *Escherichia coli* and *Helicobacter pylori*; contracting serious chronic health conditions such as chronic tuberculosis, diarrhea, and malaria; and developing certain chronic inflammatory disorders such as periodontal disease (Finch & Crimmins, 2004), which is related to increased threat sensitivity as well as anxiety and depressive disorders (Zheng et al., 2021). Population-wide vaccination availability has also changed over the past century (Gostic et al., 2016), which has greatly altered the likelihood of developing serious infections that exert persistent effects on CNS and immune system dynamics that in turn shape experiences of social safety and threat.

Like birth year, birth month is also relevant for Social Safety Theory. Seasonality affects access to healthier, non-inflammatory foods during pregnancy, which are generally more available during spring and summer months. However, seasonality also affects an infant’s risk of being exposed to pathogens that promote inflammation, which is generally more likely in autumn and winter months due to heavier rainfall and flooding (Doblhammer & Vaupel, 2001). Childhood exposure to pro-inflammatory foods and pathogens both educate the immune system and can lead to a more pro-inflammatory phenotype, making birth month important to consider.
We are not presently aware of any studies that have integrated birth cohort into research on how social stressors impact the brain or immune system to shape experiences of social safety or social threat. Age is frequently controlled for in studies of stress, health, and aging. However, testing birth cohort as a potential moderator is a very different research strategy—one that may yield new insights into previously unappreciated macro factors that shape our social lives and health.

**Culture**

As alluded to previously, each individual—and our varied social networks—are situated within different intersecting social institutions and cultures. There is an extensive literature connecting cultural norms, such as stoicism and gender norms, to health behaviors, and a growing corpus connecting similar norms to biological outcomes. In addition to differences in diet and nutrition, variation in social integration and emotional processes have been found to contribute to lower levels of inflammatory markers in Japanese adults as compared to Americans (Coe et al., 2011). Similarly, acceptance of negative emotions (characteristic of Japanese participants relative to Americans in one study) may partly explain smaller increases in IL-6 during these emotional states (Miyamoto et al., 2013). Whereas expressing anger was associated with reduced health risks (as indexed by cholesterol, blood pressure, CRP, and IL-6) in Japanese participants, the opposite was found in Americans (Kitayama et al., 2015). Although culture should not be conflated with national origin or thought of as monolithic or unchangeable (Singer et al., 2016; Shattuck, 2019), results such as these suggest that differences in behavioral and cognitive norms—certainly a part of many definitions of culture—are able to ‘get under the skin’ and affect physiological processes that are relevant for health and well-being.

Like the social institutions discussed above, cultural practices can also shape social networks, influence sources of social support, and greatly impact social safety, social threat, and health. For example, Dressler and colleagues (1986) found that support from fictive kin (ritual co-parents, or compadres) was a stronger predictor of lower blood pressure in a Mexican town than support from ‘true’ relatives, friends, or neighbors. Likewise, among the Mosuo of Southwest China, who live in both matrilineal and patrilineal groups in adjacent communities based on geography (Matthewson et al., 2016), chronic inflammation and hypertension were both lower in members of the dominant sex (Reynolds et al., 2020). More specifically, women had lower CRP levels and hypertension in matrilineal communities wherein men had higher CRP and hypertension rates and, critically, this pattern was reversed in patrilineal communities (Reynolds et al., 2020). Appreciating the role that culture plays in shaping not only social networks but also the relative perceived importance of sources of social support will be an important avenue for testing and extending Social Safety Theory.

Similar biopsychosocial effects may be induced during the process of acculturation, as individuals integrate into new cultural models. Along these lines, Gendron et al. (2020) proposed that moving to new cultural contexts should disrupt allostatic because the brain develops culturally specific models of the world based on predictive processing, as described above. Therefore, when mismatches occur between these cultural models and new sociocultural environments, allostatic load begins to accrue with potential downstream effects on inflammatory biology, stress, depression, and other sequelae (Gendron et al., 2020). This is similar to Dressler’s (2007, 2018) ‘cultural consonance’, or the degree to which individuals adhere to cultural norms of beliefs and behaviors, wherein low cultural consonance is associated with greater stress and worse health. Consistent with this formulation, high salivary cortisol and elevated Epstein–Barr virus antibodies have both been found in instances of acculturation and low cultural consonance (Decker et al., 2003; McDade et al., 2000); in contrast, seeking social support in culturally appropriate ways has been related to lower CRP levels (Dressler, Balieiro, Ribeiro, & dos Santos, 2016). Of course, social networks themselves shift during migration and acculturation, a phenomenon that requires additional study in relation to health.

Broadly speaking, the idea that culture influences how people experience the social world is not new, but we are not aware of any studies that have examined these dynamics in the context of also
assessing multi-level neural-immune interactions that have the ability to impact health or behavior. Put more simply, there are no studies on the cultural psychoneuroimmunology of social stress and health. Future research on this topic could determine the individual or interactive effects of changing cultural models and social networks on inflammatory and other health markers, and hopefully address the important question of how culture influences differences in individual and collective disease risk, leading to more precise theory, a better understanding of stress biology, and better strategies for intervening to bolster resilience, especially in cultural minority populations.

**Air pollution**

As reviewed elsewhere (Olvera Alvarez et al., 2018), ambient air pollution, including ozone, nitric oxide, sulfur dioxide, carbon monoxide, and particulate matter (PM), can affect the activity of the social signal transduction pathways underlying Social Safety Theory. These health-relevant pollutants are emitted from many common sources, including stovetops and heating units in households as well as vehicles, commercial factories, and animals in the surrounding environment (Slavich, 2020a). The most well-studied pollutant in conjunction with social drivers of health is probably PM$_{2.5}$, named so because it is ≤2.5 μm in aerodynamic diameter. Exposure to these particles has been associated with an increased risk for several inflammation-related health problems, including asthma, cardiovascular disease, stroke, cognitive impairment, depression, and suicide. These particles are especially harmful because they can be inhaled, enter the circulation, and affect brain structure, function, and plasticity by inducing local, peripheral, or systemic inflammatory events via reactive oxygen species or NF-κB, interleukin-1β (IL-1β), and TNF-α signaling (Olvera Alvarez et al., 2018).

The literature examining how social threats interact with air pollution to affect health and mortality risk is growing rapidly. In one of the largest studies on this topic so far, researchers analyzed data from more than sixty thousand American adults and found that PM$_{2.5}$ < 12 μg/m$^3$ exposure predicted a 14% increase in all-cause mortality (hazard ratio = 1.14; 95% CI [1.13, 1.14]); however, this mortality risk was significantly greater for low-SES persons and for Black participants, who are among those most likely to be discriminated against in the United States (Di et al., 2017). Documenting the magnitude of this effect, Black adults had a PM$_{2.5}$-related risk of mortality that was three times greater than that of the overall population.

Looking forward, research is needed to investigate what other types of social threats interact with air pollution to affect health and behavior, as well as precisely how these interactive effects occur. For example, does pollution exposure lower the threshold at which ACEs or adulthood life stressors induce a systemic inflammatory response, or does pollution increase the magnitude or prolong the length of the response? Along different lines, how persistent are the negative effects of air pollution in early life? Does air pollution-related priming of the social threat-induced inflammatory response persist across the lifespan despite changing environments or, alternatively, do these interactive effects diminish for those exposed to less air pollution in adulthood? Finally, what types of social-environmental circumstances or airborne pollutants are most detrimental? Given the data described above (e.g., Di et al., 2017), air pollution-mediated effects on perceptions of social safety and threat will likely be high for those living near highways or commercial zones (in urban environments) or who cook with biofuels (in rural environments), but more research is needed to clearly understand how air pollution acts both directly and indirectly to affect experiences of social safety and threat.

**Individual Difference Factors**

**Genetics and gene expression**

Turning now to individual difference factors, a variety of genetic factors and processes can influence the activity of social signal transduction pathways that link experiences of social safety and threat with health. These factors include DNA methylation, histone modification, and non-coding RNA (i.e., epigenetic modifications) as well as single nucleotide polymorphisms (SNPs), or genetic variation
at a particular genetic locus that can affect a gene’s function and the likelihood that a particular social safety or threat signal gets converted into health – and behavior-altering biology. Differences in gene expression also play a critical role. The literature on this topic is sizable and has been reviewed elsewhere (e.g., Cole, 2014; Slavich & Cole, 2013; Szyf, 2019). Here, therefore, we focus on prototypic human studies that have investigated the moderating influence of particular SNPs, given the bulk of research on this topic, and studies examining social stress-induced changes in the expression of pro-inflammatory and antiviral immune response genes that are relevant for health and behavior.

As described by Slavich and Cole (2013) and Slavich (2020a), SNPs involved in the social stress response play a key role in Social Safety Theory because, among other things, they represent a critical ‘gate’ through which social-environmental signals must travel to affect health and behavior. metaphorically speaking, when a particular genetic gate is wide open – that is, the SNP confers high transcription factor binding affinity for excitatory factors or low transcription factor binding affinity for inhibitory factors – then the likelihood that a particular social-environmental signal will impact behavior is high; conversely, when a gate is closed – that is, when a SNP confers low transcription factor binding affinity for excitatory factors or high transcription factor binding affinity for inhibitory factors – the likelihood of a particular social-environmental signal having a strong biobehavioral effect diminishes. Functional SNPs that have been found to moderate the effects that friendly and threatening social experiences have on human health and behavior involving these or other mechanisms include those in *BDNF* such as Val66Met, which codes for the brain-derived neurotrophic factor protein, and in *CRHR1* and *FKBP5*, which affect glucocorticoid receptor sensitivity and corticotropin-releasing hormone activity. Other relevant SNPs include those in the antiviral immune response genes *IFNA* and *IFNB*, which target intracellular pathogens such as viruses, and in the inflammatory cytokine genes *IL1B*, *IL6*, *IL8*, and *TNF*, which regulate immunological responses to bacteria and other extracellular pathogens.

In a prototypic example of this phenomenon, Cole et al. (2010) investigated how a functionally active regulatory SNP in the human *IL6* promoter (−174G > C, rs1800795) related to participants’ likelihood of dying from inflammation-related and all-causes of death. They found that those experiencing high social adversity who were homozygous for the GATA1-sensitive G allele died 2.8 years sooner, on average, than those with the GATA1-insensitive C allele. Moreover, this effect was found only for inflammation-related causes of death – namely, cardiovascular disease, Alzheimer’s disease, and cancer. Therefore, mortality risk attributable to inflammation-related diseases for participants reporting high social adversity was moderated by this functionally active regulatory SNP, which the authors found converted social adversity into increased SNS activity that upregulated *IL6* gene expression and increased individuals’ risk for inflammation-related causes of death. In contrast, possessing the GATA1-insensitive C allele completely mitigated the negative effect that high social threat normally has on health and mortality.

In another example focusing on a different biological system, Slavich et al. (2014) found that participants’ likelihood of developing major depressive disorder (MDD) following a major life event involving targeted rejection – that is, the intentional severing of an important social bond – was moderated by a functional SNP in the μ-opioid receptor gene (*OPRM1*, rs1799971). Specifically, G allele carriers were twice as likely to meet criteria for MDD relative to A/A homozygotes, presumably because the former exhibit less opioid receptor expression and signaling efficiency in response to socially painful experiences, such as targeted rejection. Consistent with this possibility, this particular SNP causes an amino acid change (N40D) that affects *OPRM1* expression, leading to known differences in social and physical pain sensitivity.

Given the known limitations of candidate gene studies, including the fact that they have faced reproducibility issues, these findings should be treated cautiously. As recently reviewed by Slavich et al. (2023), however, research on social experiences and gene expression has generally confirmed the involvement of these biological pathways in linking social experiences and health. In particular, socially safe environments marked by maternal warmth, stability, and predictability have been found to be consistently associated with less pro-inflammatory and more antiviral
gene expression (e.g., Chen et al., 2011; Miller et al., 2009a), whereas socially threatening environments characterized by social isolation, rejection, devaluation, and exclusion have been found to be related to more pro-inflammatory and less antiviral gene expression (Miller et al., 2009a, 2009b). What remains unknown in this body of work is the extent to which social experiences shape hour-by-hour and day-to-day changes in perceptions of social safety and threat, with downstream consequences for pro-inflammatory and antiviral gene expression. Combining genetic protocols with ecological momentary assessment methods would provide a fruitful approach for investigating these issues. Such work would enhance our understanding of the dynamic interplay of social experiences, biology, and health, and also improve our ability to identify individuals at the greatest risk of developing health problems following socially painful or threatening situations.

Sleep
Individual differences in sleep are also important to consider, given known associations between sleep and experiences of social threats, and a large body of research showing that sleep is strongly tied to health disparities (Hale et al., 2020). Sleep complaints are reported by nearly half of all adults, but serious sleep problems are especially common among individuals who have been exposed to social threats, such as interpersonal violence and aggression (Gallegos et al., 2021). In addition to activating the HPA axis and SNS, poor sleep leads to decreased antiviral gene expression and increased pro-inflammatory gene expression, which, if sustained, can increase a person’s risk for both viral infections and inflammation-related morbidity and mortality (Slavich, 2020a, 2020b, 2022). Moreover, the increase in the circulating pro-inflammatory cytokines IL-1β, IL-6, and TNF-α that occurs following poor sleep can heighten neurobiological threat sensitivity, thereby degrading perceptions of social safety and making individuals more likely to experience, and/or respond strongly to, social threat.

In a thoughtful study linking social threat, sleep, and health-relevant gene expression, Chiang et al. (2019) measured sleep duration and both pro-inflammatory and antiviral gene expression in adolescents and found that greater daily experiences of social threat – namely, arguing with parents, family members, teachers, or friends; being punished; and being insulted or threatened – interacted with shorter sleep duration to predict reduced expression of 511 genes and increased expression of 1,894 genes including several that promote inflammatory activity, such as TNF, IL1RAP, IL2RB, and IL15RA. Moreover, consistent with the social signal transduction pathways described by Social Safety Theory (Slavich, 2020a, 2022), promoter-based bioinformatics analyses revealed that these gene expression changes were mediated by increased NF-κB activity, thus implicating SNS activation in driving these differences in gene expression. The authors also went one step further and investigated the extent to which general (i.e., non-socially threatening) stress levels interacted with poor sleep to yield the same gene expression results, but they found that general assessments of stressful life events and perceived stress did not. Consistent with Social Safety Theory, therefore, the findings indicate that it is social threats in particular, rather than general life events or perceived stress in general, that interacts with sleep to affect antiviral and pro-inflammatory gene expression with downstream implications for health.

One fact that makes sleep interesting to study in this context is that associations between sleep and experiences of social safety and threat are likely bidirectional. For example, delayed sleep onset or poor sleep quality can upregulate inflammatory activity and lead to greater experiences of social threat that further disrupt sleep (e.g., poor sleep → inflammation → social threat → poor sleep). At the same time, experiences of social threat (e.g., ruminating about a hostile boss) can prevent sleep onset and/or degrade sleep quality, leading to greater inflammatory activity and experiences of social threat (e.g., social threat → poor sleep → inflammation → social threat). Given the complexity of these interrelations, future research on Social Safety Theory will greatly benefit from longitudinal studies that can discern the relative ordering of changes in these behavioral, immunologic, and cognitive processes to better understand where in the mechanistic chain interventions might be able to short-circuit the self-promoting loop linking social safety and threat, sleep, inflammation, and health.
Given the paucity of studies that include all of these factors, additional research is warranted and may yield significant benefit, given the sizable disease burden associated with experiencing life stress, chronic inflammation, and poor sleep (Fuligni et al., 2021).

**Diet**

There are several pathways by which diet also can affect experiences of social safety and threat, many of which involve inflammatory mechanisms. Overall, what a person eats impacts their health, body composition, and weight. Western, high-fat diets are associated with increased storage of adipose tissue, or fat cells. Under conditions of obesity, in turn, adipose tissue secretes high levels of pro-inflammatory cytokines in response to stress (Reilly & Saltiel, 2017), which can affect neurocognitive perceptions of threat. Additionally, the accumulation of adipose tissue can upregulate the release of pro-inflammatory cytokines through production of leptin (Iikuni et al., 2008), again having neurocognitive effects.

Diet also impacts immunologic processes that affect social safety, social threat, and health by modifying the composition of the gut microbiome. Strikingly, there are thousands of strains of bacterial microorganisms in the human gut that make up the microbiome, along with approximately 100 million neurons in the small intestine alone (Furness, 2008). Dietary patterns alter the composition of the microbiome and can influence experiences of social safety and threat by modifying the activity of the brain-gut-microbiome axis, which includes the CNS, neuroendocrine and neuroimmune systems, autonomic nervous system, enteric nervous system, and intestinal microbiota (Dinan & Cryan, 2012). Specifically, high-fat diets have been found to heighten neuroendocrine reactivity and inflammatory activity, which, in turn, induce anxiety, hypervigilance, and greater perceived social threat (Dutheil et al., 2016). Beyond these direct biological pathways, weight gain from high-fat diets and overeating – among a myriad of other factors that cause weight gain – can lead to feelings of social stigmatization and threat (Hunger et al., 2015; Tomiyama, 2014), thus promoting a feed-forward cycle of weight gain, perceived social threat, and heightened inflammatory activity.

In addition to intensifying perceptions of social threat, inflammatory activity is a cue to an organism that, because of current or perceived illness, injury, or social threats, the future is uncertain (Dantzer, 2001). Consequently, inflammation is associated with impulsivity and present-focused decision-making (Gassen et al., 2019; Shields et al., 2017). When focused on the present as opposed to the future – and especially when stressed – people tend to make food choices that are high in caloric density and immediately rewarding, but that increase the risk for future health problems driven by systemic chronic inflammation (Adam & Epel, 2007).

Just as dietary choices and eating behavior can influence perceptions of social safety, perceived social safety can impact eating behavior. For example, early life stress, including resource scarcity and unpredictability, is known to promote opportunistic eating in adults (i.e., eating in the absence of hunger; Hill et al., 2016). Unpredictability in childhood is largely determined by the reliability of adults in children’s social networks and by the safety of their environments. In one study on this topic, researchers found that children as young as seven years old typically ate homeostatically – that is, they ate more when they were hungry than when they were full. However, youth exposed to high levels of unpredictability and a lack of safety tended to eat in the absence of hunger, eating comparable amounts when they were hungry as when they were full (Leyva et al., 2020). In ancestral environments, where calories were often scarce, eating opportunistically in unpredictable and resource-scarce conditions facilitated survival. In our modern environment, however – where calories are abundant but low quality – these eating patterns likely contribute to the increasing obesity levels found among those who grow up in the context of modern resource scarcity (Nettle et al., 2017). Based on these findings, interventions aimed at increasing social safety through improving the predictability and safety of children’s environments may have a substantial positive impact on their eating behaviors, both in childhood and throughout their adult lives.
Interventions for promoting social safety may also improve the efficacy of diet interventions delivered in adulthood. An organism without social safety has limited options for how to store excess or future calories, something that is especially true in the context of human ancestral environments. Even in modern-day environments, food in the pantry does not last forever; as such, one option is to store future calories as fat to provide a safety net against starvation in case access to food becomes scarce. An organism with social safety, however, can theoretically store future calories in their social network as favors that can be called upon in dire times (Nettle et al., 2017; Neuberg & Kenrick, 2017; Pascoe & Smart Richman, 2018). Using this rationale, one potential application of Social Safety Theory lies in its ability to guide the improved effectiveness of diet interventions. Beyond impacting psychological mechanisms through which social safety can increase self-efficacy and well-being, increasing social safety could modify biological mechanisms that perpetuate links between diet and eating behaviors, adipose storage, and inflammation. Given the alarmingly high levels of obesity worldwide, and especially in western societies (Blüher, 2019), tackling this issue from the perspective of Social Safety Theory should be a top priority.

**Personality**

In addition, individual differences in personality can significantly influence how people experience the social world through multiple biobehavioral pathways. For example, individuals who are more extraverted tend to be more outgoing and motivated to seek social experiences compared to those who are less extraverted. In a study investigating emergent friendships in a cohort of students entering a new M.B.A. program, for example, incoming students who were more extraverted tended to develop more friends than their more introverted counterparts (Feiler & Kleinbaum, 2015). This personality-driven behavior could, in turn, lead to greater social safety during critical social-environmental transitions that many people find stressful.

Although personality is thought to be relatively stable (Cobb-Clark & Schurer, 2012; Sanderman & Ranchor, 1994), bidirectional relations between life experiences and personality traits likely also exist whereby significant life events, especially those occurring in early development, can influence personality and vice versa. Consistent with this possibility, Shields et al. (2016) found that changes in perceived stress levels were associated with changes in trait pessimism over five weeks in a cohort of 332 young, middle-aged, and older adults. Moreover, research that has investigated associations between personality and inflammation has revealed bidirectional relations between personality and health that are mediated through inflammatory processes (D’Acquisto, 2017; Luchetti et al., 2014; Segerstrom, 2000; Sutin et al., 2010). Indeed, effects have been found between inflammation and all Big 5 personality traits: whereas extraversion, neuroticism, and agreeableness have been related to greater inflammatory activity, openness to experience and conscientiousness have been found to be related to less inflammatory activity (Allen & Laborde, 2017; Armon et al., 2013; Chapman et al., 2011; Luchetti et al., 2014; Sutin et al., 2010). Linking this research to health, individuals high in neuroticism are known to be at greater risk of inflammation-related diseases including depression, cardiovascular disease, asthma, and irritable bowel syndrome, compared to their low-neuroticism counterparts (Lahey, 2009).

A limitation of this work is that it is difficult to interpret if neuroticism is a causal factor in these associations, as it could also be the case that a predisposition for developing inflammatory disease leads to greater neuroticism. Alternatively, it might be that individuals with elevated neuroticism develop elevated inflammation because they perceive the world as more socially threatening or because they lack social safety. Ultimately, the strength and directionality of associations between personality, inflammation, and health requires further study, and so too does the issue of how these bidirectional dynamics influence perceptions and behaviors involving social safety and social threat.

One recent study that began to untangle the causal direction of these effects examined associations between personality traits and genetic polymorphisms that are known to impact pro-inflammatory cytokine activity (Mengelkoch et al., 2022). In 213 participants, the authors found that
possessing polymorphisms that correlate with increased pro-inflammatory cytokine activity was related to being more extraverted. In a separate cross-sectional study, the researchers found that greater extraversion was related to greater stimulated pro-inflammatory cytokine release. Together, these results suggest that a heightened pro-inflammatory tendency may predispose individuals to greater extraversion, with the researchers theorizing that a heightened pro-inflammatory tendency might buffer individuals from costs associated with increased pathogen exposure inherent in traditional social interactions with others, as opposed to increased sociality being the sole causal factor leading to elevated inflammation. Given associations between elevated inflammation and loneliness and social isolation, the hypothesis that extraverted individuals are predisposed to having pro-inflammatory tendencies is one that should be tested empirically within the framework of Social Safety Theory. Importantly, this example highlights that the relation between personality and inflammation is likely bidirectional, with personality predicting behaviors that affect inflammation on the one hand, and inflammatory biology shaping aspects of personality and behavior on the other hand.

Looking forward, several pressing questions remain unanswered in this context. For example, although multiple dimensions of personality are related to inflammatory levels and likely influence experiences of social safety and threat, how social safety schemas mediate these effects has not been studied. Moreover, do personality trait levels shape peoples’ need for social safety, subjective perceptions of their own social safety, or how people react biopsychosocially to social threat? Finally, which of these associations are mediated by the immune system? Personality is a relatively understudied factor in psychoneuroimmunology that we believe should be included in studies seeking to extend Social Safety Theory, as it is likely that personality dimensions, specifically extraversion and neuroticism, moderate the impact of social experiences on social safety perceptions and schemas.

**Self-harm**

A final, high-priority individual difference factor relevant for Social Safety Theory is nonsuicidal self-injury (NSSI). NSSI involves the ‘direct, deliberate destruction of one’s own body tissue in the absence of intent to die’ and is relatively common in adolescence, with a lifetime prevalence of 23% during this developmental period (Nock, 2009, p. 78). Emotional, physical, and sexual abuse are serious social threats that have been related to NSSI (Nock, 2009; Stewart et al., 2019). Although the biobehavioral processes linking social threat and NSSI are not fully understood, inflammation has been hypothesized as a key mechanistic candidate (e.g., social threat → inflammation → NSSI; Brundin et al., 2015).

In addition to the possibility of inflammatory biology may play a role in NSSI, Slavich (2020) hypothesized that the opposite may also occur – namely, that engaging in NSSI such as cutting or burning oneself may upregulate inflammatory activity in a way that increases perceptions of social threat, thus heightening emotional distress and the likelihood of engaging in NSSI (e.g., NSSI → inflammation → social threat → NSSI). Although the biobehavioral mechanisms underlying this reverse association have not been longitudinally studied, tissue damage is known to cause the release of mitochondrial damage-associated molecular patterns (DAMPs) into the circulation, and mitochondrial DAMPs in turn initiate innate and adaptive immune responses by activating intracellular receptors (e.g., TLR9, NLRP3) or cell surface receptors (e.g., P2X7R, FPRs) on immune cells, including those that regulate neuroinflammation (e.g., microglia, astrocytes). The resulting release of inflammatory cytokines can degrade perceived social safety and increase neural sensitivity to social conflict, criticism, and exclusion, leading to greater distress and propensity to engage in NSSI (e.g., NSSI → inflammation → social threat → NSSI; Slavich et al., 2010; Zhang et al. 2010). In short, self-harm may be triggered by low perceived social safety but may also lead to such perceptions by activating social signal transduction pathways that upregulate inflammatory activity and heighten perceived social threat.
Despite the pressing need to elucidate modifiable biopsychosocial mechanisms underlying NSSI, to our knowledge, no studies have investigated how NSSI, inflammation, and experiences of social safety and threat change in concert with one another over time or, more specifically, if NSSI increases inflammatory activity in a way that modulates social cognition. It also remains unknown if just-in-time interventions that target social safety or social threat might help reduce cognitive-emotional processes that drive NSSI or that could decouple socially threatening experiences from engaging in NSSI. Given the current high prevalence of NSSI in youth, we believe that additional research into biopsychosocial processes that underlie NSSI is a high public health priority, and that using a Social Safety Theory lens for this work may be fruitful.

Measuring social safety and social threat: Laying the foundation for a field

Having explained key concepts related to social safety and threat and several key factors that may influence these constructs, we turn now to the important question of how best to assess social safety and social threat. This theory, and its scientific evaluation, are young. Therefore, it is imperative to lay a solid methodological foundation upon which to test and refine these ideas.

The first step involves measurement development and validation in each key area: perceptions, characteristics of the social situation or life event, and schemas. Measuring these three aspects of social safety and threat will provide an opportunity to begin disentangling the social situation from momentary perception from a person’s general disposition, which is important for determining what processes most strongly drive downstream effects on biology, health, and behavior. This work will also help researchers understand the dynamic interplay of social experiences, social cognition, biology, and health, as well as the extent to which perceptions and schemas change over time and what social cues most strongly induce these changes.

Some existing measures can be used to systematically assess exposure to different types of social threats over the life course. In particular, the Stress and Adversity Inventory (STRAIN) assesses lifetime stressors that possess a number of different core social-psychological characteristics, including interpersonal loss and humiliation – two key forms of social threat (Slavich & Shields, 2018; Slavich et al., 2019). A more comprehensive instrument designed to measure experiences of both social threat and safety, however, would be even more valuable insofar as it would advance our understanding of how both of these dimensions of human experience affect health and behavior.

Following the development and validation of general social safety and threat measures that assess social perceptions, experiences, and schemas, it would be useful to create modified versions to assess certain social domains, including in dyadic relationships (e.g., romantic or parental relationships), as well as in specific social environments and systems including micro- and macro-environments. These modified measures could even focus respondents on a specific interpersonal life event, such as a recent relationship break up or argument, which may help to elucidate the mechanisms underlying peoples’ interpretation of social situations, how social safety schemas are involved, and which stressor characteristics most strongly affect perceptions of social safety and threat, and their likelihood of shaping risk and resilience in the face of adversity.

It is likely that the extent to which socially safe and threatening experiences, schemas, and perceptions have implications for health partially depends on how a person values social relationships and their goals within those relationships, which are both person- and domain-dependent. One possibility is that safety and threat in domains of a person’s social life that are central to their identity may impact health and well-being the most. Similarly, because interpersonal goals both shape and are shaped by peoples’ beliefs about relationships, they likely intertwine with social schemas to affect perceptions of social safety and threat, and together create upward or downward spirals in relationship functioning via their responses and relationship behaviors (Canevello & Crocker, 2010, 2015).

Finally, two relationship beliefs relevant for Social Safety Theory that will be important to include in future research are relationship growth beliefs (Knee, 1998) and nonzero-sum beliefs (Crocker
et al., 2017). Relationship growth beliefs involve the idea that successful relationships are cultivated and developed, which can be contrasted with destiny beliefs, in which relationships are seen as stable, unlikely to change, and either meant-to-be or not (Knee, 1998). In turn, zero-sum beliefs involve the idea that relationships work in zero-sum ways – that the satisfaction of the needs and desires of one person must necessarily come at the expense of others vs. the belief that win-win solutions that are good for both people are possible (i.e., nonzero-sum beliefs; Crocker et al., 2017), which could affect the extent to which a social situation is perceived as threatening. Relationship growth beliefs may influence whether a interpersonal disagreement indicates that a relationship is doomed (high threat) or something that can be worked through and overcome (low threat), the latter of which could strengthen the relationship. (Non)zero-sum beliefs, in turn, would further encourage either a competitive or collaborative attitude to problem-solving in relationships (Crocker et al., 2017). Given the lack of scales that comprehensively assess social safety and social threat processes, much remains to be understood about how these constructs can best be assessed together.

Measuring Social Safety and Social Threat: Process Issues

The development of psychometrically sound measures is an important step toward being able to empirically evaluate Social Safety Theory. To provide more practical guidance in this pursuit, we turn now to several key considerations – namely, scope of content, developmental considerations, the need for validation in intersectionally diverse samples, and considerations for repeated measures use.

Scope of content

As described throughout this review, we conceptualize social schemas as a hierarchical construct primarily consisting of social safety and social threat, which are conceptually distinct. Both constructs are characterized by a variety of attitudes, beliefs, and expectations about the social self, world, and future, as described above. Given the natural variety of socially salient events – and consequently the attitudes, beliefs, and expectations related to these events – comprehensive measurement is critical. Therefore, ideal social safety schema measures should aim to cover a breadth of safety and threat-related attributions/beliefs (e.g., friendly vs. hostile, predictable vs. unreliable, supportive vs. critical, helpful vs. harmful, and sincere vs. manipulative) about a variety of different social contexts/events [e.g., the ability to cope with threats and conflict (i.e., social self-schemas); fundamental quality and motivations of family members, close friends, authority figures, and strangers (i.e., social world schemas); and likelihood of socially safe and threatening situations occurring in the future (i.e., social future schemas)]. Consequently, it is imperative that breadth of content not be sacrificed for high internal consistency (Clark & Watson, 2019; Cronbach & Meehl, 1955) or short measures. Should abbreviated measures be desired, they should be made after, or in tandem with, a comprehensive, gold-standard measure of social safety schemas.

Additionally, there are a number of important general recommendations for measure creation that should be heeded. Firstly, given evidence that self-report measures frequently have items that are interpreted differently by different participants (Cohen et al., 2022), instructions, item prompts, and response options must be carefully crafted to minimize measurement error. For example, double-barreled items (i.e., items that collapse one or more explicit questions into one item) should be avoided. It is also necessary to carefully select what quality will be reflected in item responses. Do item responses reflect differences in severity or frequency? Are absolute or relative scales of reference used (e.g., ‘Rarely’, ‘Sometimes’, ‘Often’ vs. ‘Never’, ‘One or two days a week’, ‘Everyday’). Critically, the ideal content to assess might vary based on context (e.g., developmental status, sociodemographic setting; see below).
Measurement invariance across development and cultural context

As discussed above, social contexts change considerably across development. Consequently, critical social events (and related schemas) will be different for someone in 3rd grade vs. college. Consequently, it is important to carefully consider how developmental context influences the ideal measurement of social safety schemas. As researchers consider how to best approach this challenge, it will be important to ask if it is even possible to develop a social safety measure that is developmentally invariant (i.e., measures social safety schemas the same way regardless of one’s developmental context) or if separate, developmentally specific measures are needed. Relatedly, given that social safety schemas are posited to be influenced by cultural norms and messaging, it is possible that ideal measures of social safety schemas could differ as a function of sociodemographic characteristics or cultural contexts. Along these lines, it will be important to evaluate if social safety measures are invariant across racial and ethnic groups, geographically disparate regions, socio-economic statuses, religious affiliations, sexual orientations, or gender identities (see Moriarity et al., 2022).

Considerations for repeated measures use

Finally, given its etiological nature, carefully testing Social Safety Theory will require longitudinal study designs. It is critical that this point be considered in the construction social safety schema measures. Many self-report questionnaires orient participants to a certain time frame (e.g., ‘How many times have you felt X in the past week’). Chronological orientation in a social safety measure should correspond to the theorized temporal stability of social safety schemas (e.g., whether it is more trait – or state-like in nature). Further, it is important for chronological orientation to complement the time between measurements in a longitudinal study. For example, it is not ideal to ask participants to complete a measure every week if the measure asks about experiences over the last month because the period reflected in the answers will overlap between measurements. Also, because it is unlikely that all hypotheses will have the same ideal time lags between measurements (i.e., temporal specificity; Moriarity & Alloy, 2021), the development of social safety measures with options for several time lags would maximize correspondence between measurement and study design, likely reducing measurement error in future work.

Interventions for promoting social safety and reducing social threat

In this last section, we briefly consider how Social Safety Theory can inform interventions for reducing social threat and enhancing individual and collective resilience. To this end, Slavich (2020a) posited that individuals are embedded within a series of social safety circles, with the individual at the center and extending out to family, school/work, neighborhood and the broader community, society, and world (see Figure 3). Each circle independently and interactively contributes to an individual’s sense of social safety and threat. Accordingly, interventions can be developed and deployed to target these different levels, such as individual psychosocial interventions that target maladaptive social cognitions, behaviors, and interpersonal patterns that degrade social safety, or school-based interventions that reduce bullying, social aggression, and exclusion, or increase social cohesion and empathy for others.

Table 3 provides a summary of interventions – at the individual, family, school, and community/society level – that may bolster social safety and reduce social threat. Note that many of these examples represent modifications that could be done to existing interventions, as informed by Social Safety Theory. For example, variants of cognitive behavior therapy (CBT) have been developed and shown to be effective for a wide variety of mental health problems, including depression (Beck et al., 1979), anxiety disorders (Barlow, 2021), eating disorders (Fairburn, 2008), substance use (Beck, Wright, et al., 2011), insomnia (Perlis et al., 2008), personality disorders (Beck et al., 2015; Linehan, 1993), bipolar disorder (Otto et al., 2008), and schizophrenia (Beck, Rector, et al., 2011). Common across these forms of CBT is the focus on teaching clients the skills to identify and modify negative patterns of thinking and maladaptive patterns of behavior that contribute to their specific symptoms.
Given its flexibility, CBT could be readily adapted to focus specifically on social safety and threat as key content dimensions. With regards to cognitive work, a Social Safety Theory-informed CBT could emphasize those cognitions that are most related to social safety and threat. For example, an individual could be taught to recognize their social safety-related core beliefs (e.g., ‘I’m unlovable,’ ‘People cannot be trusted’) and related dysfunctional attitudes (e.g., ‘If I show my true self, I’ll get rejected’) that contribute to a heightened sense of social threat and decreased experiences of social safety. Existing CBT tools and techniques, such as thought records (Beck, 2011), could be easily tailored to help individuals modify these maladaptive cognitions and develop healthier patterns of thinking. Similarly, with regards to behavioral work, Social Safety Theory-informed CBT could shift the focus toward encouraging behaviors that specifically involve fostering actual and perceived social safety, such as calling or visiting a loved one, scheduling more quality time with a good friend, writing a gratitude or love letter to a partner, and making a better effort to socialize and develop friendships with neighbors.

In addition to the core cognitive and behavioral work, CBT therapists often provide training in other beneficial social skills, including interpersonal problem solving and assertiveness. Informed by Social Safety Theory, interpersonal problem-solving strategies could be taught to address
ongoing relationship problems to reduce the likelihood of current and future disagreements, conflict, and interpersonal stressors from occurring. In addition, individuals with a more passive communication or interpersonal style could benefit from assertiveness training to enhance their confidence and sense of agency, which could ultimately lead to more fulfilling social interactions and relationships. In summary, CBT provides an empirically validated and highly flexible framework that could be tailored for targeted interventions aimed at decreasing actual and perceived social threat, and increasing social safety. Given evidence that targeting social processes is highly therapeutic (Lipsitz & Markowitz, 2013), our contention is that psychotherapies may be most efficient and effective if they specifically modify social cognitions and behaviors involving the dimensions

**Table 3. Individual and collective strategies for promoting social safety and reducing social threat.**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Possible Mechanism of Action</th>
<th>Example Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive behavior therapy</td>
<td>Increase social engagement and reduce social threat-related thoughts and beliefs</td>
<td>Clark &amp; Beck, 1999; DeRubeis et al., 2010; Hofmann &amp; Otto, 2017; Shields, Spahr, &amp; Slavich, 2020</td>
</tr>
<tr>
<td>Acceptance and commitment therapy</td>
<td>Develop cognitive flexibility to notice, accept, and embrace past social threats</td>
<td>Hayes, Strosahl, &amp; Wilson, 2009</td>
</tr>
<tr>
<td>Mindfulness meditation</td>
<td>Bring awareness to the present, which is typically absent of social threat</td>
<td>Black &amp; Slavich, 2016; Creswell, Pacilio, Lindsay, &amp; Brown, 2014; Dahl et al., 2015</td>
</tr>
<tr>
<td>Loving-kindness and compassion meditation</td>
<td>Enhance positive emotional states of kindness and compassion toward others</td>
<td>Boellinghaus et al., 2014; Hofmann, Grossman, &amp; Hinton, 2011</td>
</tr>
<tr>
<td>Forgiveness therapy</td>
<td>Reduce negative thoughts and emotions induced by offenders/aggressors</td>
<td>Akhtar &amp; Barlow, 2018; Worthington, 2013</td>
</tr>
<tr>
<td>Self-distancing</td>
<td>Reduce the negativity of past aversive events by increasing distance/perspective</td>
<td>Kross &amp; Ayduk, 2017</td>
</tr>
<tr>
<td>Stress mindset interventions</td>
<td>View social threats as enhancing rather than debilitating experiences</td>
<td>Crum et al., 2017</td>
</tr>
<tr>
<td>Acetaminophen administration</td>
<td>Reduce neurobiological signaling that subserves social pain and rejection</td>
<td>DeWall et al., 2010; Slavich et al., 2019</td>
</tr>
<tr>
<td>Safety learning</td>
<td>Remodel neural networks underlying anxiety, fear, and threat</td>
<td>Meyer et al., 2019</td>
</tr>
<tr>
<td><strong>Family</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurturant parenting training</td>
<td>Strengthen familial relationships</td>
<td>Miller et al., 2014</td>
</tr>
<tr>
<td>Family cohesion training</td>
<td>Promote social caring, interpersonal connection, and family cohesion</td>
<td>Jalali et al., 2016; King et al., 2019</td>
</tr>
<tr>
<td><strong>School</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity safety interventions</td>
<td>Foster sense that one’s identity and values are accepted and welcome</td>
<td>Goyer et al., 2019</td>
</tr>
<tr>
<td>Social belonging interventions</td>
<td>Increase social connection, cohesion, belonging, affiliation, and inclusion; reappraise</td>
<td>Allen, Vella-Brodrick, &amp; Waters, 2016; Borman, Rozek, &amp; Pyne, 2019; Patton et al., 2006; Walton &amp; Cohen, 2011</td>
</tr>
<tr>
<td>Mere belonging interventions</td>
<td>Enhance social connection and belonging with minimal cues and socially shared goals</td>
<td>Gehlbach et al., 2016; Walton et al., 2012</td>
</tr>
<tr>
<td><strong>Community/Society</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduce bias</td>
<td>Decrease instances of prejudice and discrimination</td>
<td>Eberhardt, 2019</td>
</tr>
<tr>
<td>Prevent bullying</td>
<td>Reduce experiences of social aggression and exclusion</td>
<td>Williams &amp; Nida, 2014</td>
</tr>
<tr>
<td>Promote social connection</td>
<td>Decrease loneliness and increase social interaction and belonging</td>
<td>Holt-Lunstad et al., 2017</td>
</tr>
<tr>
<td>Build empathy</td>
<td>Promote social connection, caring, and empathy toward others</td>
<td>Zaki, 2019</td>
</tr>
</tbody>
</table>

G. M. SLAVICH ET AL.
of social safety and threat. Indeed, as described by Slavich (2020a), disruptions in social cognition and behavior are a core feature of all major forms of psychopathology, making the focus on disrupted social cognition and behavior potentially critical for therapeutic success.

These are just a few of the many interventions that may help promote social safety and reduce social threat. In addition to these individual interventions are numerous family, school, and community-level interventions that could also help improve the social environment within which individuals are embedded (see Table 3). Looking forward, there are a number of important questions that future social safety-related intervention research should consider. In particular, which intervention (or combination of interventions), at which level, is most effective for promoting social safety and reducing social threat? Moreover, do some subgroups of individuals benefit more from some interventions than others? In addition, research is needed to tailor social safety-related interventions to each client’s specific characteristics and presenting problem (i.e., precision medicine), and to test whether interventions can be modified to prevent the onset of social stress-related disorders before they occur (Cuijpers et al., 2021). Only by pursuing this intervention work will we realize the most important benefits of Social Safety Theory.

Conclusion

In conclusion, although we have learned a great deal about the psychobiology of stress and health over the past few decades, there is surprisingly little agreement about which characteristics of positive and negative social experiences are most impactful. In addition, we still lack a complete understanding of how such experiences impact the brain and body to shape health, development, and behavior across the life course. The advent of affordable methods for assessing stress-relevant biology is changing all of this, though, and if we use the resulting data wisely, the result could be an exciting new understanding of stress, health, and resilience that is biologically grounded and evolutionarily informed. This has been our goal in describing and extending Social Safety Theory herein – namely, to take what we already know about stress biology and use this knowledge to identify the social experiences that should be most important to focus on given the highly evolved regulatory logic of the human brain and immune system. We believe that such an understanding can help reveal the biopsychosocial roots of lifelong health disparities and, equally important, point researchers toward new treatment targets for reducing disease risk and achieving health equity. Given the enormous disease burden caused by stress and adversity – especially in vulnerable and disadvantaged populations – we simply cannot think of a more important issue to tackle.

Acknowledgements

We thank the handling editor and two anonymous reviewers whose comments helped to greatly improve this article.

Disclosure statement

The authors declare no conflict of interest with respect to this work.

Funding

George Slavich, Lydia Roos, Daniel Moriarity, Summer Mengelkoch, and Jenna Alley were supported by grant #OPR21101 from the California Governor’s Office of Planning and Research/California Initiative to Advance Precision Medicine. Daniel Moriarity was additionally supported by National Institutes of Health (NIH) National Research Service Award F32 MH130149, and Christian Webb was partially supported by NIH grant R01 MH116969, NIH grant R01 AT011002, the Tommy Fuss Fund, and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. These organizations had no role in writing this article or in deciding to submit it for publication.
References


