Disruption, Self-Presentation, and Defensive Tactics at the Threshold of Learning

Alex Gillespie[1]

Abstract

Disruptive experiences are opportunities for learning, yet, people often resist them. This tendency is evident in individual experience, organizational behavior, and denialist discourses. Research has been hampered by conceptualizing this defensiveness in terms of unconscious defense mechanisms or underlying cognitive processes. In contrast, I conceptualize defensiveness in terms of observable defensive tactics, namely, the actions and utterances that are used to resist disruptive meanings. I introduce the analogy of the semantic immune system to conceptualize three layers of defensive tactics: avoiding, delegitimizing, and limiting the impact of disruptive meanings. Defensive tactics are cultural–historical creations that, like the immune system, have adapted over time to neutralize disruptive meanings. I use this tripartite conceptualization to review the fragmented literature on defensive tactics. The observability of these tactics gives centrality to the audience who either calls out or does not call out the use of defensive tactics—questioning or implicitly supporting the legitimacy of the defended views. The vigilance of the audience pushes the tactics toward increasingly subtle forms that seek to pass undetected. Reconceptualizing defensiveness in terms of observable tactics reveals the importance of the audience and opens these tactics up to empirical research, calling upon researchers to identify the increasingly subtle ways in which learning through dialogue is inhibited.

Keywords

defensive tactics, denialism, defense mechanisms, audience, learning

Disruptive experiences are opportunities for learning, yet, people often resist them. This goes to the heart of learning, where the disjunction between expectation and experience provides an opportunity to refine one's expectations and thus avoid future disruptions. Defensiveness toward disruptive meanings can undermine this potential for learning. What motivates such defensiveness? How is it actually performed? And what are its limiting conditions?

I conceptualize defensiveness in terms of defensive tactics that are observable in talk and action, and an audience that judges the reasonableness of what is being said and done. To date, research has focused on the unconscious and cognitive sources of defensiveness. From the initial defense mechanisms proposed by Freud to more recent conceptualizations of cognitive dissonance, researchers have aimed to understand what happens intrapsychologically. In contrast to these views, which, despite their differences, conceptualize defensive talk and action are mere epiphenomena, I argue that observable defensive tactics are constitutive.

Conceptualizing defensiveness in terms of defensive tactics is a cross-disciplinary contribution to general psychology. It combines ideas from learning theories, cultural psychology, conversation analysis, and organizational psychology to address limitations in how development, or learning, through social feedback is conceptualized. Moreover, the conceptualization of defensive tactics contributes at multiple psychological scales, being applicable at the individual, interactional, group, and organizational levels of analysis.

The article begins by reviewing the role of disruptive experiences in theories of learning proposed by Dewey, Piaget, Bateson, and Friston. These individualistic and cognitive theories are expanded in three ways. First, I argue that other people are a source of disruptive experiences, and the gaze of the other motivates defensiveness. Second, I argue that defensiveness manifests itself in defensive tactics that are empirically observable in talk and action. Third, because defensive tactics are observable, they are judged by

[1]The London School of Economics and Political Science, London, UK

Corresponding Author:
Alex Gillespie, Department of Psychological and Behavioural Science, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.
Email: a.t.gillespie@lse.ac.uk

audiences. When audiences "call out" defensive tactics, it pushes them toward increasingly subtle forms. Indeed, across these sociohistorical tactics, there is an arms race between the audiences who are on the look-out for defensiveness, and the defenders who self-present as being reasonable. I introduce a novel metaphor of the semantic immune system to provide an integrative analogy. In doing so, I delineate three layers of defensive tactics: avoiding, delegitimizing, and limiting the impact of disruptive experiences. I then use this tripartite distinction to provide an integrative review and typology of defensive tactics.

Learning and Rupture

Theories of individual and organizational learning have emphasized the importance of disruptive experiences that challenge or contradict people's expectations. Disruptive experiences, in contrast to mundane or routine experiences, require adjustment; they necessitate a change to the guiding representation, self-concept, or course of action. This basic idea forms a common thread in the learning theories developed by Dewey, Piaget, Bateson, and more recently Friston.

The basic idea of learning through rupture was introduced in Dewey's (1896) naturalistic account of thinking as part of perception–action–consequence loops. Insofar as perception, action, and consequence are in alignment, humans are experientially embedded or absorbed in their activity (Csikszentmihalyi, 1997). But, when the consequence of an act does not conform to expectations (e.g., the child reaches for a flame expecting to find a play-thing), the perception–action–consequence loop breaks down. Cognition then arises in the attempt to reconstruct perception and response so as to achieve the desired outcome. "Life," Dewey (1922, p. 179) wrote, "is interruptions and recoveries." Thus, according to Dewey, disruptive experiences not only instigate cognition, but also stimulate reorganization of the cognitive system.

Piaget (1977) refined the definition of disruptive experience by distinguishing between assimilation and accommodation. For Piaget, assimilation is the integration of external information into preexisting cognitive structures or ways of thinking. Exchanging pleasantries, routine coordination, or receiving an answer to a mundane question entails the assimilation of information without the need to adjust the schema into which the information is assimilated. However, new meanings can sometimes only be internalized, or understood, if there is a change to the preexisting schema. For example, when a child learns about the conservation of liquid, their worldview is altered. Equally, when an adult receives negative feedback on their job performance, they might have to make an accommodation in their self-concept. In Piaget's view, accommodation is any modification of the assimilatory schema or semantic worldview brought about by the information being assimilated (Block, 1982). Building on Piaget, I define a disruptive experience as any experience that conflicts with expectation to the extent that an accommodation must be made.

Bateson's (1972) distinction between single- and double-loop learning is similar to Piaget's distinction between assimilation and accommodation. Single-loop learning occurs when there is adjustment of action toward an unquestioned goal and there is no learning at the level of goals, beliefs, or learning itself (e.g., assimilating information). Double-loop learning entails realizing that single-loop learning occurs within the context of a particular goal, belief, or value. This realization enables a person to modify their goal, belief, or value, or select an entirely new goal, belief, or value (i.e., accommodate a change in the assimilatory schema). However, Bateson goes beyond Piaget to conceptualize a third level of learning. Triple-loop learning occurs when the person acknowledges that double-loop learning itself occurs within specific contexts (e.g., a set of assumptions, habits, potentials) and that they can learn about the context of those contexts (e.g., by questioning one's assumptions or becoming aware of one's resistance to learning). For Bateson, as with Dewey and Piaget, disruptive experience is the mechanism that pushes learning to higher levels. Contradictions between perceptions, actions, and consequences lead one to question the goal of those actions (double-loop learning), and contradictions within the set of possible goals leads one to learn about learning (triple-loop learning). But contradictions themselves are not enough, what is required for learning, is the metalogical realization that one's own goals, beliefs and assumptions are fallible (Sammut & Gaskell, 2010). An aspect of Bateson's (1972, p. 303) thinking that is particularly relevant for the present article is his idea of "loopholes"—ways of thinking that allow people to live with contradictions and thus inhibit the push to a higher level of learning. The defensive tactics, that I conceptualize later, explain how these loopholes operate.

More recently, Friston's (2010) unifying theory of active inference reiterates the basic idea of learning through disruptive experiences. Learning, he argues, brings the environment and the living system's representation of that environment into equilibrium. The environment, via sensation, shapes the internal representation; the internal representation, via action, shapes the material and social environment (Constant et al., 2019). Life is the creation of action-oriented representations of the environment, that through loops of sensation and action, aims to minimize surprise or disruptive experiences. Disruption can lead to a reduction of future surprise either through learning (i.e., updating the internal representation) or through action (i.e., avoiding the source of surprise). Friston's theory of active inference overemphasizes cognition, neglecting the extent to which culture and the external environment channel
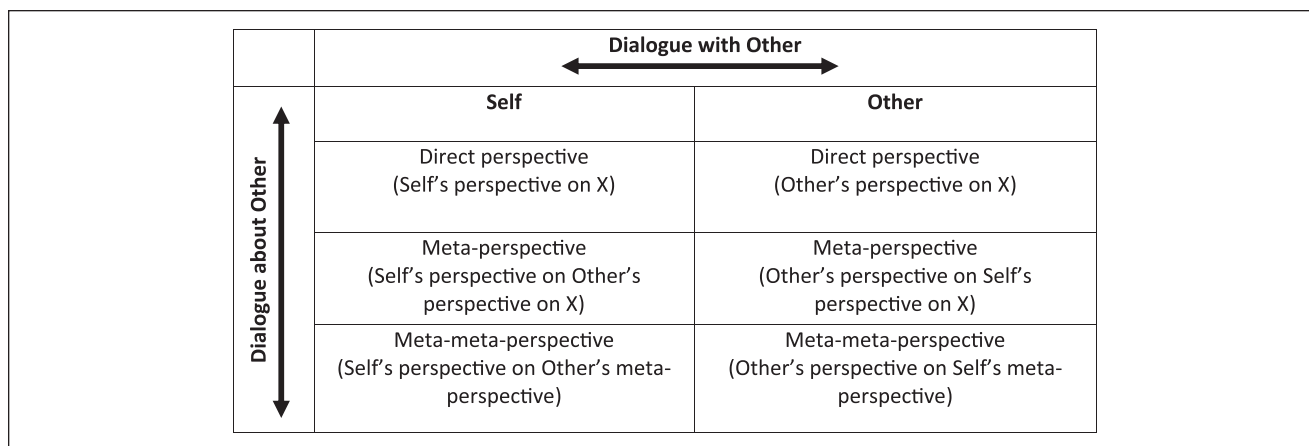
Figure 1. Intersubjective relations.

human thought and action (Clark, 2018). Moreover, Friston seems to portray people as naïve scientists and fails to fully account for why and how people might resist learning from disruptive experiences.

These foundational theories of learning share the idea that disruptive experiences are a dynamic driver of learning. To take this basic idea further, I argue that this rupture-learning model needs to be complemented by an understanding of why individuals often resist the opportunity to learn from disruptive experiences. To address this issue, the basic rupture-learning idea needs to be extended to consider the role of other people in causing disruptions and motivating defensiveness, the specific tactics through which defensiveness is achieved, and the role of the audience in calling out defensive tactics.

The Rupture of the Other

The focus of Dewey's, Bateson's, and Friston's theories is on the isolated actor earnestly trying to understand the material world. But, as Piaget (1932) recognized, contradictions often arise in social interaction. Children must understand the conservation of liquid not because it is true, but, because it is the only way to share juice between peers without getting criticized (Psaltis & Duveen, 2007). The social world (i.e., other people) is both a source of disruption and a motivator of defensiveness.

To understand the role of others in creating disruptions, I conceptualize the social world as an intersubjective world—a matrix of perspectives (shaped by interests, perceptions, cognitions, identities, cultures, etc.) engaged in cooperative and competitive social acts (Tomasello et al., 2005). We can distinguish three layers of intersubjectivity that occur both between and within people (Gillespie & Cornish, 2010; Laing et al., 1966). First, direct perspectives refer to how Self and Other directly perceive any object "X." Second, meta-perspectives refer to how Self and Other think each

other perceives "X." Third, meta-meta-perspectives refer to how Self and Other understand (usually vaguely) how each other perceives their view of "X." Figure 1 depicts these three layers of intersubjectivity.

This intersubjective framework reveals two types of social interaction between Self and Other, each with two types of potential disruption. First, there is prototypical social interaction, namely, dialogue with Other (face-to-face or mediated, verbal or nonverbal). Such interactions can present ruptures of disagreement (Self and Other have different direct perspectives on X) or misunderstanding (Self's direct perspective does not match Other's meta-perspective on Self's perspective). Second, there is internalized social interaction, namely, dialogue about Other (where people talk or think about the perspective of Other). These interactions can produce ruptures of perceived disagreement (Self's direct perspective does not match Self's meta-perspective on Other) and perceived misunderstanding (Self's direct perspective does not match Self's meta-meta-perspective on Other). These internalized intersubjective interactions are evident in the dialogicality of talk (Bakhtin, 1986). People do not speak only about their own views; they also speak about the views of others, especially those they disagree with. Speakers are constantly positioning themselves within an argumentative landscape and thus is replete with tensions between direct perspectives and meta-perspectives (Holt, 2000).

Unlike contradictory experiences in the material world (e.g., the pain of the flame mistaken for a toy, not believing in the conservation of liquid), however, contradictions in the social world are not resolved by capitulating to a single truth. The truth of the social world is that it contains many perspectives on what is true; to understand the social world is to acknowledge this diversity of perspectives. In short, the contradictions of social life are not only "out there," they necessarily become part of the self (Marková, 2016). Our identity, our very sense of self, is created in the

intersubjective web of perspectives—in how other people see us and how we think other people see us (Goffman, 1959; Mead, 1934).

If an experience or idea is dissonant with one's self-presentation or with the norms associated with a positive sense of self, the disapproving gaze of the audience motivates the individual to resolve that dissonance (Aronson, 2019; Stone & Cooper, 2001). However, this gaze can have a perverse effect on learning; the individual forgoes the opportunity for long-term learning in favor of a short-term gain by promoting the impression that they have nothing to learn. Understanding how this works requires the concept of defensive tactics.

Defensive Tactics

Although Bateson and Friston provide some insight into the loopholes and failures of learning, the literatures on defense mechanisms and cognitive dissonance can broaden their conceptualizations. Investigating defensive tactics instead of defense mechanisms, however, will, I argue, provide a more empirically tractable route to understanding defensiveness at the threshold of learning.

The psychoanalytic conception of defense mechanisms was developed in detail by Anna Freud (1937). Defense mechanisms describe how people unconsciously handle clashes between their desires and reality. These mechanisms include denial (pushing discomforting thoughts out of experience); isolation (minimizing associative connections to unacceptable or dangerous ideas); displacement (redirecting emotion at a more acceptable target); reaction formation (covering up an unacceptable impulse by behaving in the opposite manner); and repression (suppressing impulses that are not personally or socially acceptable). Experimental research has supported several of these defense mechanisms (Baumeister et al., 1998), showing them to be particularly active in response to threat (Cramer, 2012). Defense mechanisms are not pathological: privileging one's desires over reality may lead to denial and even delusion, but disregarding one's desires in the face of reality can lead to despondency and undermine resilience (Cramer, 2000).

The social psychology literature conceptualizes defensiveness in terms of cognitive dissonance—the discomfort people feel in response to conflicting information (Festinger, 1957). Researchers have identified several cognitive processes, besides adjusting one's beliefs (i.e., learning), that reduce the dissonance and feelings of discomfort. Selective exposure refers to the tendency of individuals to seek affirming information and avoid disconfirming information (Hart et al., 2009). Motivated reasoning refers to the tendency to engage in nonrational thinking that reinforces pre-existing beliefs (Kunda, 1990). In cases of identity-threat, self-affirmation and outgroup derogation can also be observed (Glasford et al., 2009).

One interesting development in the research on cognitive defenses is inoculation theory (McGuire, 1961). Based on the metaphor of inoculation against disease, this theory suggests that we could:

> develop the resistance to persuasion of a person raised in an ideologically aseptic environment by pre-exposing him to weakened forms of the counter-arguments, or to some other belief-threatening material strong enough to stimulate, but not so strong as to overcome, his belief defences. (McGuire & Papageorgis, 1962, p. 25)

This theory has grown and produced a substantial body of evidence since McGuire's original formulation. Research has revealed the cognitive (e.g., accessibility, rationality), affective (e.g., threat, anger), and individual (e.g., personality, self-efficacy) processes that underpin inoculation (Compton & Pfau, 2005).

Although the research on defense mechanisms, cognitive dissonance, and inoculation has yielded valuable insights into the underlying mechanisms, researchers have treated the defensive behavior as a secondary manifestation of those mechanisms. In his reanalysis of some of Freud's case studies, Billig (1999) argued that we should look for defense mechanisms, such as repression, "in" what someone says and does, rather than "behind" what someone says and does. In the talk within the therapeutic encounter, people "push away disturbing thoughts in much the same way as we avoid troublesome topics in conversation" (Billig, 1999, p. 38). In light of this, Billig (1999) argued for "a surface psychology which puts language at the centre of human thinking" (p. 38). In effect, Billig's critique (which he extends to cognitive psychology) was that it is more empirical to study language and behavior itself, rather than to speculate about the processes behind these ostensibly superficial manifestations.

I build on Billig's (1999) paradigmatic analysis to recast questions of defensiveness in terms of dialogical interaction and thus pose new questions: What do rationalization, denial, isolation, and displacement look like in dialogue? What are the practical and conversational tactics used to avoid topics that cause cognitive dissonance? Might inoculation be conceptualized as providing people with resources to silence disruptive meanings? What effect does the audience calling out defensive tactics have on the effectiveness of those tactics? Overall, the idea is that these ostensibly surface manifestations are not mere epiphenomena of underlying processes. Rather, the success or failure of defensiveness rests upon the defensive tactics used and specifically how their use is viewed by the audience. Defensive tactics are defined as talk and action that circumvent disruptive meanings under the oversight of one or more audiences. The literature on defensive tactics is broad and fragmented, cutting across traditional disciplinary boundaries and levels of analysis.

Organizational researchers have examined defense tactics to understand problematic organizational cultures that appear unable to learn (Hede & Bovey, 2001; Homburg & Fürst, 2007; Westrum, 2004). Argyris (1990, 2003) argued that double-loop learning in organizations is inhibited by defensive routines, for example, procrastinating, crowding out disruptive topics, asking pointless questions, creating barriers to bad news, not passing on bad news, maintaining taboo topics, engaging in strategic ineffectiveness, and sending mixed messages (e.g., espousing being a "learning organization," but failing to act on disruptive learning opportunities).

Conversation analysts studying dialogue with Other have examined the discursive tactics people use to both introduce disruptive meanings and silence them in face-to-face interactions. For example, when broaching difficult topics, people tend to use hedges, contradictions, formality, impersonality, and humor (Chen, 2001). When trying to silence someone speaking up, they may speak over that person, stigmatize them, or undermine their motive (Tholander, 2011). People complaining, raising concerns, or even asking questions can threaten accommodation, or double-loop learning, and thus stimulate defensiveness.

Sociocultural psychologists studying intrapersonal dialogue and dialogue about Other have examined how people talk about and resist others' points of view (Kadianaki, 2014; Moore et al., 2011; Sammut et al., 2013). First, reported speech gives voice to the other ("they think"); then, the speaker responds ("but. . ."). Such dialogism within a conversational turn is evident when police defend themselves against accusations of racism (Morant & Edwards, 2011) and in argumentative positioning around immigration (Gillespie et al., 2012; Tsirogianni & Andreouli, 2011). People use semantic barriers, such as dichotomies, taboos, and stigma, to neutralize threatening meanings (Gillespie, 2008; Moscovici, 1976).

A complete conceptualization of defensive tactics needs to include organizational, interpersonal, and intrapersonal levels of analysis, as these levels coexist. Organizational defensiveness invariably has interpersonal and intrapersonal aspects; equally, any intrapersonal defensiveness can spill over into interpersonal and organizational contexts. Moreover, not only are similar tactics evident in these different domains and levels, but the central role of the audience is also evident throughout.

Audience and Appearing Reasonable

Given that defensiveness is motivated by the gaze of others and defensive tactics are observable, it follows that defensive tactics invariably have an audience. At a minimum, people are their own audiences, but usually there is also someone being communicated with (e.g., a friend, someone posting on social media, an interviewer), and sometimes there are also third parties (e.g., observers of the talk or social media post, or the recording of the interview).

Putting the audience at the center of defensive tactics helps to address what Cohen (2001, pp. 5–6) has called the "denial paradox," namely that to resist disruptive meanings, one must on some level understand the implications of those meanings (i.e., if one is defensive one cannot be in denial). Across human history, were the leaders of highly stratified societies troubled by, and thus defensive toward, class conflict or inequality? Although they doubtless were defensive toward other phenomena, the problems of class and inequality require a modern audience. Reconceptualizing defensiveness as an interactional performance avoids the denial paradox. When person B raises a problem for person A, this threatens person A's identity which in turn motivates A to engage in face-saving tactics. Person B may say that person A is "in denial" about the problem; but, the repressed problem is not in the mind of person A, it is in the mind of person B. The problem is, however, so vivid to person B, they think it must be also in person A's mind, albeit buried.

This is not to deny intrapsychological processes. The audience, I suggest, is integral to creating more psychological and delusional beliefs. For example, an alcoholic whose family or friends do not speak up will quite likely be deluded about their alcoholism. If nobody mentions the problem, then, in terms of social reality, it is not a problem. To the extent that defensiveness operates in the twilight of awareness where phenomena are ambiguous, people look to each other as anchor-points in constituting social reality (Bem, 1972). Therefore, well-meaning family and friends can even become complicit in sustaining delusional beliefs.

Defensive tactics which appear reasonable have the potential to make the audience complicit in legitimating the defensive definition of the situation. Outright exclusion, denial, and denigration are uncommon because they can backfire (i.e., appearing defensive can give credibility to the disruptive meaning). Goffman (1959) described this as a kind of arms race in which the presenter develops tactics to impress the audience, while the audience becomes increasingly skilled at detecting self-presentational tactics. Although defensive tactics can certainly succeed even when the audience recognizes them as such (e.g., in hierarchical or autocratic situations), they are most effective when they go undetected. For example, a government minister facing a scandal may set up a public inquiry to cloak inaction (e.g., via lenient investigators, limited remit or powers, delays) in the progressive discourse of transparency and learning. Successfully concealed defensive tactics that are not publicly called out can provide a legitimacy boost; they create a norm of acceptance that can even bolster the defender's self-image as being a reasonable person (Cohen, 2001). Thus, the goal is rarely defensive tactics *per se*, but rather successfully implemented stealthy tactics. At the same time, the audience is not passive, and they will learn to recognize and distrust

novel tactics (Grenier et al., 2012). When the audience then calls out such tactics it can (but does not necessarily) undermine their effectiveness (Schmid & Betsch, 2019), thus pushing the tactics toward ever more subtle forms.

Going further, one can argue that the audience pushes the dialogue toward being reasonable. Developmental research has shown that rationality is closely bound to social interaction (Carpendale & Müller, 2014; Piaget, 1932). Indeed, being rational is arguably that which works to convince others through argumentation (Habermas, 1981). Reason is "providing reasons" and is thus a form of public justification that withstands scrutiny. Without an audience, it is easy to be unreasonable; it is easier to dismiss someone's disruptive idea in dialogue about them than in dialogue with them. This is because most audiences demand reasonableness and can call out denials, distortions, and omissions of inconvenient details. In this way, other people play an important role not only as a source of disruptive experiences or ideas, but also as detectors of defensive tactics and even the upholders of the standards of rationality. Audiences tends to be more aware of defensive tactics than the defenders realize (Gibb, 1961); in speaking up the audience limits our more elementary impulses toward denial, pressing us toward, if not a more honest engagement with disruptive experiences, at least more stealthy defensive tactics.

The Semantic Immune System

People often resist disruptive meanings and the opportunities for learning they present in an effort at self-presentation. However, for it to succeed, the defensive tactic needs to avoid being called out. Fortunately for defenders, there are a wide variety of tactics to choose from. To conceptualize the different types of defensive tactics, I propose to expand the root metaphor of inoculation theory (McGuire, 1961) into the concept of the semantic immune system.

The immune system is a layered defense system (Herzenberg & Herzenberg, 1989), reflecting phylogenetic and ontogenetic development as the system moves from generic to increasingly specific responses. First, the innate immune system provides generic defenses that ignore the specific features of any pathogen. The most basic defense layer is made up of physical barriers (e.g., skin, respiratory tract, gastrointestinal tract), secretions (e.g., mucous, saliva, sweat), and general immune responses (e.g., inflammation, marking cells for destruction). Second, the adaptive immune system learns and provides defenses that are tailored to the specifics of each pathogen. Inoculation works at this adaptive layer: providing a small dose of a pathogen stimulates the adaptive immune system to develop defenses that are tailored to that pathogen.

On a general level, the immune system has evolved to address a problem that is common to all bounded systems. Although boundaries preserve the integrity of systems, those systems also need to transact with an environment (e.g., animals have mouths, doors have walls). The problem is that some transactions can undermine the integrity of the system and thus triage is necessary (e.g., the digestive system separates desirable from undesirable elements, locks on doors separate owners and guests from intruders). The immune system provides a highly advanced solution to this general problem, making it a powerful analogy for understanding comparable domains. For example, the immune system analogy has been influential in the development of modern computer security policies and procedures (Somayaji et al., 1997) and the filtering of digital information before it gets to users (Chao & Forrest, 2003). Here, I apply the analogy to the domain of humans transacting with disruptive ideas.

I conceptualize the semantic immune system as composed of three sequential layers of defensive tactics (avoiding, delegitimizing, and limiting) that people use to defend against disruptive meanings. The phrase "disruptive meaning" contains no value judgment; it refers in a technical sense to meanings that are disruptive to preexisting meanings because they require double-loop learning (accommodation). The first layer, akin to the skin, focuses on avoiding encounters with disruptive meanings (e.g., ignoring, delaying). The second layer, akin to the generic innate immune system, focuses on generic delegitimizing tactics (e.g., stigmatization, distrust). The third layer, akin to the adaptive immune system, focuses on limiting the impact of engaging with the semantic content of the disruptive meaning (e.g., rationalization, isolation). As with the immune system, each layer entails an increasingly specific response.

Each layer of defense can also be distinguished in terms of the Self–Other–Object triangle that underpins many classic social psychology theories (Zittoun et al., 2007). This triangle aims to identify the minimal unit of social interaction. Avoiding operates on the self (i.e., preventing Self encountering disruptive meanings); delegitimizing operates on the other (i.e., attacking Other as valid source of meaning); and limiting operates on the object (i.e., belittling the implications of the disruptive meaning).

In the following three sections, I present a novel review of defensive tactics conceptualized in terms of the semantic immune system. I identified the tactics through an analysis of diverse bodies of literature, including research on organizations, conversations, and intrapsychological defense processes. The review began with a search for literature on "defenses" and "denial," but I expanded the review beyond these search terms to examine literature citing or being cited in identified texts. Across the literature, I focused on tactics (i.e., practical and communicative acts) and excluded consideration of purely intrapsychological processes (e.g., repression, cognitive dissonance). All the tactics identified are acts that are publicly evident and observable by audiences, which, as I have argued, is integral to their functioning.

Table 1. Avoiding tactics.

Tactic	Definition	Illustration
Excluding	Not giving the source of disruptive meaning a space to speak or be heard.	For example, excluding women from speaking (Houston & Kramarae, 1991), not letting stakeholders participate in decision-making (Habermas, 1989), not having processes for hearing dissent, complaints, or whistleblowing (Homburg & Fürst, 2007), sacking employees who raise problems (Vuori & Huy, 2016), using gatekeepers to filter disruptions (Janis, 1972), and selectively avoiding sources of disruptive meaning (Hart et al., 2009).
Ignoring	Not responding to an encountered disruptive meaning, or acting as if it does not exist.	For example, selectively attending to the least disruptive aspects of a problem (Chen, 2001; Cramer, 2012; Stich & Wagner, 2012), ignoring an issue or source (Cramer, 2000; DeFrancisco, 1991; Graziano et al., 1980; Sweeny & Shepperd, 2007), talking about groups without giving voice to their interests (O'Doherty & Lecouteur, 2007), crowding out disruptive meanings in talk (Tholander, 2011), filling agendas and time with nondisruptive topics (Argyris, 1990), and not asking questions that might uncover disruptive meanings (Gouran et al., 1986).
Discouraging	Creating an atmosphere, cultivating a personality, or demanding burdens of proof that inhibit disruptive meanings from being voiced or developed.	For example, "temperamental" leaders (Vuori & Huy, 2016), autocratic leadership styles (Noort et al., 2019), and hierarchies themselves (Reader et al., 2007) can inhibit people from speaking up. Burdening includes: requirements to find more evidence, reformat, resubmit, restate, or represent the disruptive meaning; setting criteria that are unachievable (Argyris, 1990); and unnecessarily burdensome procedures for complaining or whistleblowing (Homburg & Fürst, 2007).
Denying	Overtly disagreeing with the disruptive meaning without any rationale.	In contrast to ignoring (the absence of response), denying is a response that refuses to acknowledge the disruptive meaning. Denying is particularly common among children (Cramer, 2012), but is also evident among adults (Brown & Starkey, 2000; Napolitano, 2018; Stich & Wagner, 2012; Sweeny & Shepperd, 2007), organizations, and governments (Cohen, 2001). This differs from denialism (e.g., of the Holocaust, HIV, vaccines), which entails intricate webs of rationalization (Mathis, 2006; McKee & Diethelm, 2010).
Distracting	Raising a separate issue to shift the focus of attention away from the disruptive meaning.	This can include idealizing the self (Cramer, 2015; Stich & Wagner, 2012), emphasizing the positive qualities of the self (e.g., "working hard"), or being overly optimistic about the future (Weick & Sutcliffe, 2003). However, it can also include raising unexpected issues, or "red herrings," that aim to derail the focus of attention (McKee & Diethelm, 2010), and "whatabouting," which involves pointing to an ostensibly comparable problem and implying double standards (Headley, 2015).
Deflecting	Deflecting the disruptive meaning toward another person, department, place, or time.	For example, denying responsibility (Alicke & Sedikides, 2009; Stich & Wagner, 2012), blaming regulations and procedures (Parker et al., 2006; Westrum, 2004), blaming others (Baumeister et al., 1998; Hudson, 2003; Joffe, 1999; Tholander, 2011), or delaying due to competing demands, immediate time pressures, and seeking out more information (e.g., via inquiry, working group, or just waiting and seeing) are used to justify procrastination (Argyris, 1990; Sweeny & Shepperd, 2007).

Avoiding Tactics

Like the skin, avoiding tactics prevent people from encountering potentially disruptive semantic content. Avoidance is an elementary response to negative experiences (Alicke & Sedikides, 2009). Just as people avoid the burning sensation of a flame, they also try to avoid disruptive meanings. Avoiding tactics can be applied generically to any disruptive meaning independent of its semantic content. These tactics range from physical movement (avoiding places, groups) and selective information exposure (avoiding websites, news stories) to distraction tactics (entertainment, idealizing self, red herrings) and outright denial. Each avoiding tactic focuses on keeping the self away from disruptive meanings. Table 1 provides an overview of avoiding tactics.

Power is particularly salient within avoiding tactics. People with power often employ personal secretaries or other gatekeepers to triage potential encounters. Furthermore, people can be reluctant to speak up to those in power (Reader et al., 2007), especially if there is a norm of consent (Noort et al., 2019). Power is also evident in dialogues about the Other: within the ingroup, without the gaze of the outgroup, there is huge leeway in how the Other is represented (Glăveanu et al., 2018). Although very effective, these avoiding tactics risk creating information siloes and distorted perceptions (Janis, 1972; Quattrociocchi et al., 2016).

Delegitimizing Tactics

The second layer of defense, like the white blood cells in our immune system, provides a rapid but fairly generic attack. Delegitimizing tactics focus not on the disruptive meaning itself, but instead on its source—"killing the messenger" (Westrum, 2004)—and undermining its epistemic authority

Table 2. Delegitimizing tactics.

Tactic	Definition	Illustration
Relativizing:	Attempting to neutralize the source by arguing that it is "just their opinion."	This tactic has unwittingly been aided by postmodernist discourses (Cohen, 2001) that give legitimacy to alternative facts (Cooke, 2017) and definitions ("violence is a subjective term"; McKee, 2006). This tactic can include the manufacture of misinformation (Cook et al., 2017), which destabilizes disruptive meanings by making them contestable and doubtable (McKee & Diethelm, 2010) within a climate of uncertainty (agnotology; Proctor, 1995). Relativizing weakens the disruptive meaning so that it can be "brushed off" with phrases such as "who cares?" (Sammut et al., 2013).
Stereotyping:	Casting the source (individual, group, or idea) as a predefined type that is already known and thus does not have anything valuable to contribute.	For example, the disruption is attributed to "the usual suspects," "only" a known person (Argyris, 1990). This tactic also includes describing people as "ignorant" (Sammut & Sartawi, 2012), uneducated, or nonexpert (Batel & Castro, 2009). In more extreme forms, it entails essentializing (Kadianaki & Andreouli, 2017; Raudsepp & Wagner, 2011) and dehumanizing (Diniz et al., 2019) people such that their perspective is considered not worth taking.
Stigmatizing:	Stigmatizing anyone who espouses the disruptive idea.	Such stigma means not only that the disruptive idea can be easily disregarded, but also that any semantic engagement with the disruptive idea (within the ingroup) becomes taboo. These tactics can include devaluing (Stich & Wagner, 2012) and ridiculing (Houston & Kramarae, 1991) people and ideas. Transfers of meaning also serve to stigmatize the source (Moscovici, 1976), for example, by referring to refugees as "illegal immigrants" (O'Doherty & Lecouteur, 2007) or claiming that someone like Hitler held the same belief (Harris et al., 2012).
Distrusting:	Attributing an ulterior motive to the source such that the disruptive meaning is framed as prejudiced, insincere, or manipulative.	For example, framing the source as only being interested in money or sex (Gillespie, 2008), promoting private interests (Jovchelovitch, 2007), or being biased toward another party (Tholander, 2011). Distrust is used in political campaigns to hinder the public from engaging with the substance of opponents (e.g., "crooked Hilary"; Oborne & Roberts, 2017) or an enemy (e.g., "axis of evil"; Aune, 2003). Denialism often utilizes conspiracy theories about ulterior motives in the medical establishment, peer-review, and grant funding to discount scientific research (McKee & Diethelm, 2010).

and legitimacy (Fricker, 2007; Jovchelovitch & Priego-Hernandez, 2015; Sammut & Sartawi, 2012). Delegitimizing tactics are less fixed and generic than avoiding tactics because they are tailored to the source, but rarely tailored to the disruptive meaning itself. Table 2 presents an overview of delegitimizing tactics.

Power operates in the second layer in terms by determining who has the legitimacy not only to speak but also to be heard. This is where epistemic injustice (Fricker, 2007) occurs, for example, when prejudice degrades the credibility of a voice, person, or group to the extent that the content of what is said loses credence and their epistemic authority to be heard collapses (Fricker, 2007). Systems for creating (and by implication denying) legitimacy include professional associations, educational attainments, institutional roles, and increasingly social media stature (e.g., marked by views, likes, and followers). Although legitimacy has always been subject to manipulation (e.g., paid-for or coerced crowds), online legitimacy has proven strikingly easy to manipulate (Yang et al., 2019).

Limiting Tactics

The third layer of defensive tactics, like the adaptive immune system, is most tailored to the specifics of the disruptive meaning. Limiting tactics acknowledge, to some extent, the disruptive meaning and work to minimize its impact. These tactics are more specific than either avoiding or delegitimizing tactics which are relatively generic responses. Because these tactics engage with the disruptive meaning itself, these tactics merge into substantial engagement (i.e., rational debate) with the disruptive meaning (Table 3).

Power operates in the limiting layer of defenses as rhetoric, persuasion, and ideology. The history of rationalizations and reactionary counter-arguments provides resources (Hirschman, 1991) and even inoculations to neutralize disruptive meanings as familiar tropes (McGuire, 1961). Emotional appeals and metaphors are used to realign disruptive meanings in terms of primary dichotomies (e.g., good/bad, true/false) and thus ideologically color the entire semantic field (Marková, 2016). Rationalizing, fudging, and placating obfuscate contradictions between what is believed, what is said, and what is done, leading the unwatchful eye to overlook a minefield of questions.

Evaluation

Using the immune system as an analogy to conceptualize defensive tactics brings together a previously diverse set of

Table 3. Limiting tactics.

Tactic	Definition	Illustration
Isolating:	Presenting the disruptive meaning as isolated, atypical, peculiar, no longer relevant or subjective.	For example, by emphasizing that the disruptive meaning is temporally distant (e.g., "in the past," "one-off," "lessons learned"; Barreiro et al., 2017; Weick & Sutcliffe, 2003), held by a minority (e.g., "rogue," "bad apples"; Hudson, 2003), or in someone's mind (e.g., "they think," "their perception"; Gillespie, 2008; Kus et al., 2013; Papastamou, 1986). This can lead to isolated attempts to resolve, placate, gag, or even buy-off problems, while overlooking the systemic issue (Conlon & Murray, 1996; Sweeny & Shepperd, 2007; Westrum, 2004).
Dichotomizing:	Creating rigid oppositions based on primitive binaries (e.g., good/bad, corrupt/ virtuous, rational/ irrational; Marková, 2016).	This process binds the disruptive meaning into a system of mutually exclusive oppositions that inhibits the possibility of thinking between the binaries. For example, the opposition between East/West has curtailed immigrants' ability to integrate ideas and experiences relating to their home and host countries (Andreouli, 2013), and the opposition between nutritious/un-nutritious has blocked meat-eaters' engagement with the idea of vegetarianism (Panagiotou & Kadianaki, 2019). "Black and White" thinking is also evident in denialist discourses, wherein any imperfection in knowledge becomes a reason to dismiss the evidence in its entirety (Mathis, 2006).
Fudging:	Using ambiguous language to avoid lying or drawing attention to a disruptive meaning.	For example, in the Challenger disaster, engineers, feeling under pressure to support the launch, said that they did not have enough information to "absolutely assure" that nothing would go wrong (Gouran et al., 1986, p. 130). Fudging is often indicated by imprecise designators (e.g., "this," "that," "it") and verbal hedges (e.g., "sort of," "a bit," "maybe"; Danescu-Niculescu-Mizil et al., 2013; Morand, 2014; White, 2003). This includes the use of phrases such as "collateral damage" and "transfer of populations" in the context of war (Cohen, 2001).
Rationalizing:	Talking or thinking about the disruptive meaning in such a way as to diminish its implications (or implicatory denial; Cohen, 2001).	For example, creating excuses (e.g., "difficult cases"; Alicke & Sedikides, 2009; Weick & Sutcliffe, 2003), explanations (rationalizing the smell of gas as mosquito spray; Weick, 2010), putting something "in context" (Conlon & Murray, 1996; Davidow, 2003; Westrum, 2004), making unacceptable behavior seem normal (Brown & Starkey, 2000; Homburg & Fürst, 2007; Stich & Wagner, 2012), arguing the issue is secondary or tangential (Stich & Wagner, 2012), believing that change is "pointless" (Argyris, 1990) or may even have a perverse effect (Hirschman, 1991), and appealing to other priorities (e.g., health & safety, protecting the environment, efficiency, rules; Argyris, 1990; Westrum, 2004).
Splitting	Paying lip service to a disruptive meaning in one context, but not following through in other contexts (Argyris, 1990).	Being contradictory (e.g., between what is said and done, or what is done or said in different contexts) solves the short-term problem of engaging with a disruptive meaning, while neutralizing the long-term impact (DeFrancisco, 1991). On a broader scale, this tactic allows for hypocrisy, with self-narratives that conform to expectations, despite deviations in behavior (Batson et al., 1999; Valdesolo & DeSteno, 2007).

defensive tactics into a model of how defensiveness is achieved. The analogy provides three mappings, or propositions, that can stimulate future research. First, it suggests that each layer of defense is logically secondary to the preceding layer (although the layers usually operate simultaneously, as with the actual immune system). For example, if an avoiding tactic is successful, then a delegitimizing tactic may not be necessary. Second, it suggests that each layer provides an increasingly specific response. For example, avoiding tactics are applicable to most disruptive meanings regardless of their source, but limiting tactics are more adapted to the disruptive meaning. Third, just as the immune system enables us to live in a world full of pathogens, the semantic immune system enables us to live in a world full of disruptive meanings, without succumbing to those meanings. Understanding that

other people have different points of view is necessary for human cooperation (Tomasello et al., 2005). The semantic immune system, arguably, enables people to simultaneously be aware of other, potentially disruptive points of view, while also resisting their implications.

To illustrate the semantic immune system, consider the antisuffragette movement in the United Kingdom in the early 20th century. Avoidance in the form of exclusion was evident in women not being able to vote, having limited access to the public sphere, and the incarceration of women like Emmeline Pankhurst. When they could not be avoided or silenced, then there were attacks on the suffragette's legitimacy via stereotyping and stigmatizing. Words and images depicted these women as "spinsters" being "mad unladylike monsters" characterized by "sexual deviance" and "masculine insecurity"

(Cooper-Cunningham, 2019, p. 394). As the movement grew, such attempts at delegitimizing became untenable. Accordingly, the antisuffragettes increasingly had to engage with the content of the suffragettes' arguments to limit their impact. For example, Lord Curzon (1912, p. 1) dichotomized men and women and their respective roles, arguing that women's suffrage would "break up the harmony of the home" and take women away from their "proper sphere and highest duty, which is maternity." Combining delegitimizing with rationalization he continued that women don't have "the calmness of temperament or the balance of mind" to "exercise a weighty judgment"; their "emotional excitement" would make politics "wholly uncertain." The three layers of the semantic immune system, with their increasingly specific responses, enabled antisuffragettes to live in a social world where suffragette ideas were in ascendance by minimizing the disruptive potential of these ideas while trying to maintain a public semblance of reasonableness.

Although the analogy of the immune system reveals patterns in human defensiveness, it oversimplifies human's defensiveness in four ways. First, disruptive meanings are different from pathogens. Although the double-loop change or accommodation (e.g., change in identity or plans) instigated by disruptive meanings can be painful and costly, it fosters adaptation and learning. Accordingly, being influenced by a disruptive idea is not the same as getting sick, in that, it is not always undesirable. Indeed, the institutions of education and public debate are directed toward creating such influence.

Second, disruptive meanings are resources for understanding other people and groups via a meta-perspective. Just like the adaptive immune system has a memory system for pathogens, arguably the semantic immune system stores representations of disruptive meanings (as weak "straw man" arguments). However, while the adaptive immune system aims to recognize and resist pathogens, the semantic immune system uses these memories as resources to understand others (i.e., they are the building blocks of meta-perspectives). Arguably, one of the functions of the semantic immune system, in contrast to the actual immune system, is to enable these meta-perspectives to coexist with the self's actual beliefs (i.e., direct perspectives).

Third, the immune system does not have an audience who judges whether a defense is reasonable thus, potentially, interfering with the effectiveness of the defense. The immune system is effective if it kills pathogens, regardless of the method. The semantic immune system, however, is constrained in using underhand methods to silence disruptive meanings. For example, the suffragettes turned attempts to silence the movement, such as removing demonstrators and incarcerating activists, into publicity (Cooper-Cunningham, 2019). When one's defensive tactics are called out, it can create the impression that the disruptive experience has "hit a nerve." The perception that one is being defensive can make one appear weak or lacking insight, create distrust, and undermine legitimacy.

Finally, the semantic immune system is continually influenced by other people, who have vested interests in whether people engage with this or that disruptive meaning. For example, gagging clauses and nondisclosure agreements make other people avoid a disruptive meaning. Equally, leaking documents and whistleblowing are attempts to circumvent such avoiding tactics. In terms of delegitimizing tactics, stigmatization and stoking distrust are commonly used to encourage a given group to resist influence. Arguably, one of the reasons for representing outgroups as untrustworthy is that it encourages ingroup members to dismiss the outgroup's ideas outright (Gillespie, 2007). Finally, in terms of limiting tactics, preemptive confessions, reinforcing dichotomies, prepackaged rationalizations, and rehearsed counter-arguments are widely used to protect audiences from disruptive meanings (Compton & Pfau, 2005).

Discussion

Humans' tendency toward a self-serving view of the world is forever challenged, not only by the material world, but also by the "exasperating otherness of others" (Hirschman, 1991, p. ix). Although the theories of Galileo, Darwin, and Freud fundamentally decentered humans on an intellectual level, the threat of decentering exists in every social encounter. To stabilize one's universe of meaning in the midst of so many alternatives, humans have developed subtle and powerful tactics that enable engagement without learning. The contribution of the present article has been to bring together hitherto separate strands of research on defensive tactics within the novel framework of the semantic immune system. This model makes the social world, especially audiences, central in generating disruptive experiences, motivating defensive tactics, and judging the reasonableness of what is said and done.

Learning from disruptive experiences is neither inevitable nor impossible; while there is a tendency toward defensiveness, that tendency is also limited. The concept of the semantic immune system helps to explain how learning is limited by defensive tactics, how using these defensive tactics is curtailed by the audience, and thus how people adapt to a world of alternative and potentially challenging points of view.

The core problem motivating the current analysis is the fact that disruptive experiences do not simply lead to learning. Double- or triple-loop learning is often unsettling and even painful (Bateson, 1972), as it entails subordinating one's representations to the environment (Piaget, 1977). Rather, it is tempting to do the opposite—to subordinate the environment to one's representations—and wallow in feelings of mastery, albeit illusory. Examining defensive tactics has revealed the specific actions and utterances that limit and circumvent learning, explaining how contradictions, which are routinely conceptualized to be a driver of learning (Friston, 2010), can be neutralized.

Why don't people simply choose illusion and dismiss all disruptive experiences? Why engage in these subtle defensive tactics? In short, what limits defensiveness itself? Defensive tactics, I have argued, are limited by the audience. The audience, when empowered, demands reasonableness, thus pushing defensive tactics via self-presentation (Goffman, 1959) toward communicative rationality (Habermas, 1981). However, the audience is only likely to have this key role in contexts where debate and critical feedback is enabled and valued. A question for future research is how disempowering voice, whether in small groups or society, interacts with the calling out of defensive tactics.

Is the semantic immune system more active, or aggressive, among some individuals? Although my focus has been on the tactics themselves, there is a connection to the well-established literature on the need for cognitive closure (Webster & Kruglanski, 1994) that requires examination. Future research should examine: do people who are high on need for closure use more defensive tactics? And, do people with different need for closure profiles tend toward distinctive defensive tactics? Addressing these questions might provide observable indicators of need for closure profiles.

Although the defensive tactics are used by individuals they exist as potential resources independent of any one individual (Gillespie & Zittoun, 2010). Social groups borrow, use, reuse, and modify defensive tactics when defending their own representations of the world. Representations that are challenged by alternatives (e.g., groups in conflict or facing a collective risk) may cultivate, refine and circulate defensive tactics that group members can use to neutralize the established semantic threat (Avraamidou & Psaltis, 2019; Joffe, 1999). Thus, the defensive tactics that comprise the semantic immune system, although instantiated through individual action, are sociocultural artifacts that develop through history independently of individuals. Future research could trace the sociohistorical development of the semantic immune system and the diverse tactics used in different cultures and at different times.

The semantic immune system also provides insight into how people live in a world of alternative points of view without experiencing cognitive dissonance (Panagiotou & Kadianaki, 2019). Disruptive ideas lurk in conversations and newspaper headlines. Moreover, awareness of other people's potentially disruptive perspectives is necessary for social coordination (Tomasello et al., 2005). Defensive tactics are the means through which the disruptive potential of these meta-perspectives. But, defensive tactics are rarely perfect. Merely talking about alternatives creates semantic contact, which in turn opens the door to the possibility of questioning one's own point of view (Coudin, 2013; O'Dwyer et al., 2016). Even if the representation of the other's perspective is a simplification, it is the first step in acknowledging that perspective and initiating dialogue (Psaltis, 2016). Thus by virtue of enabling people to live with disruptive meanings, defensive tactics somewhat paradoxically can open the door to learning (Batel & Castro, 2009).

The focus on the role of the audience creates new avenues for research. In Bateson's (1972) terminology, disruptive meanings imply a double-loop adjustment. These double-loop accommodations, however, are constrained at the third level (i.e., defensive tactics). The semantic immune system theorizes the constraints on double-loop learning in terms of defensive tactics; the push toward triple-loop learning is the audience's awareness of those tactics. Research has already shown that bringing visibility to rhetorical tactics reduces the persuasiveness of science denialism arguments (Cook et al., 2017; Schmid & Betsch, 2019). In the same vein, does having one's own defensive tactics called out facilitate learning? More generally, how do audiences enforce standards of reasonableness? And, are defensive tactics that are not noticed or called out by an audience more effective than those that are? If so, in the arms race of self-presentation (i.e., circumventing disruptive meanings without appearing defensive), which tactics are currently succeeding? Such questions prompt research to get ahead of the development of new defensive tactics, drawing attention to them wherever they arise.

Defensiveness, I have argued, can be studied as empirically observable tactics that manifest in talk and action. The observability of these tactics gives a central place to the audience. The gaze of the audience can threaten one's identity thus motivating defensive tactics. But, the gaze of the audience, if it detects defensive tactics, can also undermine the efficacy of those tactics. It is crucial to better understand how people avoid, delegitimize, and limit disruptive meanings under the watchful gaze of various audiences. Identifying these increasingly subtle defensive tactics can foster triple-loop learning.

ORCID iD

Alex Gillespie [iD] https://orcid.org/0000-0002-0162-1269

References

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*(1), 1–48.

Andreouli, E. (2013). Identity and acculturation: The case of naturalised citizens in Britain. *Culture & Psychology*, *19*(2), 165–183. https://doi.org/10.1177/1354067X13478984

Argyris, C. (1990). *Overcoming organizational defenses*. Allyn & Bacon.

Argyris, C. (2003). A life full of learning. *Organization Studies*, *24*(7),1178–1192.https://doi.org/10.1177/01708406030247009

Aronson, E. (2019). Dissonance, hypocrisy, and the self-concept. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed., pp. 141–157). American Psychological Association.

Aune, J. A. (2003). The argument from evil in the rhetoric of reaction. *Rhetoric & Public Affairs*, *6*(3), 518–522. https://doi.org/10.1353/rap.2003.0052

Avraamidou, M., & Psaltis, C. (2019). Blocking the solution: Social representations of threats and (non) dialogue with alternative representations in Greek-Cypriot newspapers during peace negotiations. *Journal for the Theory of Social Behaviour*, *49*, 460–479. https://doi.org/10.1111/jtsb.12222

Bakhtin, M. (1986). *Speech genres & other late essays*. University of Texas Press.

Barreiro, A., Wainryb, C., & Carretero, M. (2017). Power struggles in the remembering of historical intergroup conflict: hegemonic and counter-narratives about the Argentine "Conquest of the Desert". In C. Psaltis, M. Carretero, & S. Cehajic-Clancy (Eds.), *History education and conflict transformation* (pp. 125–145). Palgrave Macmillan.

Batel, S., & Castro, P. (2009). A social representations approach to the communication between different spheres: An analysis of the impacts of two discursive formats. *Journal for the Theory of Social Behaviour*, *39*(4), 415–433. https://doi.org/10.1111/j.1468-5914.2009.00412.x

Bateson, G. (1972). *Steps to an ecology of mind*. Ballantine Books.

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*(3), 525–537.

Baumeister, R. F., Dale, K., & Sommer, K. L. (1998). Freudian defense mechanisms and empirical findings in modern social psychology: Reaction formation, projection, displacement, undoing, isolation, sublimation, and denial. *Journal of Personality*, *66*(6), 1081–1124.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp 1–62). Academic Press.

Billig, M. (1999). *Freudian repression: Conversation creating the unconscious*. Cambridge University Press.

Block, J. (1982). Assimilation, accommodation, and the dynamics of personality development. *Child Development*, *53*(2), 281–295. https://doi.org/10.2307/1128971

Brown, A. D., & Starkey, K. (2000). Organizational identity and learning: A psychodynamic perspective. *Academy of Management Review*, *25*(1), 102–120.

Carpendale, J. I. M., & Müller, U. (2014). Social interaction and the development of rationality and morality: An introduction. In J. I. M. Carpendale & U. Müller (Eds.), *Social interaction and the development of knowledge* (pp. 1–18). Psychology Press.

Chao, D. L., & Forrest, S. (2003). Information immune systems. *Genetic Programming and Evolvable Machines*, *4*(4), 311–331.

Chen, R. (2001). Self-politeness: A proposal. *Journal of Pragmatics*, *33*(1), 87–106.

Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, *17*(3), 521–534. https://doi.org/10.1007/s11097-017-9525-z

Cohen, S. (2001). *States of denial: Knowing about atrocities and suffering*. John Wiley.

Compton, J. A., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association*, *29*(1), 97–146.

Conlon, D. E., & Murray, N. M. (1996). Customer perceptions of corporate responses to product complaints: The role of explanations. *Academy of Management Journal*, *39*(4), 1040–1056.

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., & Friston, K. (2019). Regimes of expectations: An active inference model of social conformity and human decision making. *Frontiers in Psychology*, *10*, Article 679. https://doi.org/10.3389/fpsyg.2019.00679

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, *12*(5), Article e0175799. https://doi.org/10.1371/journal.pone.0175799

Cooke, N. A. (2017). Post truth, truthiness, and alternative facts: Information behavior and critical information consumption for a new age. *The Library Quarterly*, *87*(3), 211–221. https://doi.org/10.1086/692298

Cooper-Cunningham, D. (2019). Seeing (in)security, gender and silencing: Posters in and about the British women's suffrage movement. *International Feminist Journal of Politics*, *21*(3), 383–408. https://doi.org/10.1080/14616742.2018.1561203

Coudin, G. (2013). The breakdown of the hegemonic representation of madness in Africa. *Journal for the Theory of Social Behaviour*, *43*(1), 23–44. https://doi.org/10.1111/jtsb.12001

Cramer, P. (2000). Defense mechanisms in psychology today: Further processes for adaptation. *The American Psychologist*, *55*(6), 637–646. https://doi.org/10.1037/0003-066X.55.6.637

Cramer, P. (2012). *The development of defense mechanisms: Theory, research, and assessment*. Springer Science & Business Media.

Cramer, P. (2015). Defense mechanisms: 40 years of empirical research. *Journal of Personality Assessment*, *97*(2), 114–122. https://doi.org/10.1080/00223891.2014.947997

Csikszentmihalyi, M. (1997). *Creativity: Flow and the psychology of discovery and invention*. Harper Perennial.

Curzon, G. (1912). *Lord Curzon's fifteen good reasons against the grant of female suffrage* (No. 5). National League for Opposing Women's Suffrage.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational*

Linguistics (pp. 1250–1259). Association for Computational Linguistics.

Davidow, M. (2003). Organizational responses to customer complaints: What works and what doesn't. *Journal of Service Research*, *5*(3), 225–250.

DeFrancisco, V. L. (1991). The sounds of silence: How men silence women in marital relations. *Discourse & Society*, *2*(4), 413–423. https://doi.org/10.1177/0957926591002004003

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*, 357–370.

Dewey, J. (1922). *Human nature and conduct*. Henry Holt.

Diniz, E., Bernardes, S. F., & Castro, P. (2019). Self- and other-dehumanization processes in health-related contexts: A critical review of the literature. *Review of General Psychology*, *23*(4), 475–495. https://doi.org/10.1177/1089268019880889

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Freud, A. (1937). *The ego and the mechanisms of defence*. Hogarth Press.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Gibb, J. R. (1961). Defensive communication. *Journal of Communication*, *11*(3), 141–148.

Gillespie, A. (2007). The intersubjective dynamics of trust, distrust and manipulation. In I. Markova & A. Gillespie (Eds.), *Trust and distrust: Sociocultural perspectives* (pp. 131–152). Information Age Publishing.

Gillespie, A. (2008). Social representations, alternative representations and semantic barriers. *Journal for the Theory of Social Behaviour*, *38*(4), 375–391. https://doi.org/10.1111/j.1468-5914.2008.00376.x

Gillespie, A., & Cornish, F. (2010). Intersubjectivity: Towards a dialogical analysis. *Journal for the Theory of Social Behaviour*, *40*(1), 19–46. https://doi.org/10.1111/j.1468-5914.2009.00419.x

Gillespie, A., Kadianaki, I., & O'Sullivan-Lago, R. (2012). Encountering alterity: Geographic and semantic movements. In J. Valsiner (Ed.), *The Oxford handbook of culture and psychology* (pp. 695–709). Oxford University Press.

Gillespie, A., & Zittoun, T. (2010). Using resources: Conceptualizing the mediation and reflective use of tools and signs. *Culture & Psychology*, *16*(1), 37–62. https://doi.org/10.1177/1354067X09344888

Glasford, D. E., Dovidio, J. F., & Pratto, F. (2009). I continue to feel so good about us: In-group identification and the use of social identity—Enhancing strategies to reduce intragroup dissonance. *Personality and Social Psychology Bulletin*, *35*(4), 415–427. https://doi.org/10.1177/0146167208329216

Glăveanu, V. P., de Saint-Laurent, C., & Literat, I. (2018). Making sense of refugees online: Perspective taking, political imagination, and internet memes. *American Behavioral Scientist*, *62*(4), 440–457. https://doi.org/10.1177/0002764218765060

Goffman, E. (1959). *The presentation of self in everyday life*. Penguin.

Gouran, D. S., Hirokawa, R. Y., & Martz, A. E. (1986). A critical analysis of factors related to decisional processes involved in the challenger disaster. *Central States Speech Journal*, *37*(3), 118–135. https://doi.org/10.1080/10510978609368212

Graziano, W. G., Brothen, T., & Berscheid, E. (1980). Attention, attraction, and individual differences in reaction to criticism. *Journal of Personality and Social Psychology*, *38*(2), 193–202.

Grenier, J., Pomeroy, B., & Reffett, A. (2012). Speak up or shut up? The moderating role of credibility on auditor remedial defense tactics. *Auditing: A Journal of Practice & Theory*, *31*(4), 65–83.

Habermas, J. (1981). *The theory of communicative action: Reason and the rationalization of society*. Polity Press.

Habermas, J. (1989). *The structural transformation of the public sphere*. Polity Press.

Harris, A. J., Hsu, A. S., & Madsen, J. K. (2012). Because Hitler did it! Quantitative tests of Bayesian argumentation using ad hominem. *Thinking & Reasoning*, *18*(3), 311–343.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, *135*(4), 555–588.

Headley, J. (2015). Challenging the EU's claim to moral authority: Russian talk of "double standards." *Asia Europe Journal*, *13*(3), 297–307. https://doi.org/10.1007/s10308-015-0417-y

Hede, A., & Bovey, W. H. (2001). Resistance to organisational change: The role of defence mechanisms. *Journal of Managerial Psychology*, *16*(7), 534–548. https://doi.org/10.1108/EUM0000000006166

Herzenberg, L. A., & Herzenberg, L. A. (1989). Toward a layered immune system. *Cell*, *59*(6), 953–954.

Hirschman, A. O. (1991). *The rhetoric of reaction*. Harvard University Press.

Holt, E. (2000). Reporting and reacting: Concurrent responses to reported speech. *Research on Language and Social Interaction*, *33*(4), 425–454.

Homburg, C., & Fürst, A. (2007). See no evil, hear no evil, speak no evil: A study of defensive organizational behavior towards customer complaints. *Journal of the Academy of Marketing Science*, *35*(4), 523–536.

Houston, M., & Kramarae, C. (1991). Speaking from silence: Methods of silencing and of resistance. *Discourse & Society*, *2*(4), 387–399. https://doi.org/10.1177/0957926591002004001

Hudson, P. (2003). Applying the lessons of high risk industries to health care. *BMJ Quality & Safety*, *12*(Suppl. 1), i7–i12.

Janis, I. L. (1972). *Victims of groupthink*. Houghton Mifflin.

Joffe, H. (1999). *Risk and "the other."* Cambridge University Press.

Jovchelovitch, S. (2007). *Knowledge in context: Representations, community, and culture*. Routledge.

Jovchelovitch, S., & Priego-Hernandez, J. (2015). Cognitive polyphasia, knowledge encounters and public spheres. In G. Sammut, E. Andreouli, G. Gaskell, & J. Valsiner (Eds.), *The Cambridge handbook of social representations* (pp. 163–178). Cambridge University Press.

Kadianaki, I. (2014). The transformative effects of stigma: Coping strategies as meaning-making efforts for immigrants living in Greece. *Journal of Community & Applied Social Psychology*, *24*(2), 125–138. https://doi.org/10.1002/casp.2157

Kadianaki, I., & Andreouli, E. (2017). Essentialism in social representations of citizenship: An analysis of Greeks' and migrants' discourse. *Political Psychology*, *38*(5), 833–848.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kus, L., Liu, J., & Ward, C. (2013). Relative deprivation versus system justification: Polemical social representations and identity positioning in a post-Soviet society. *European Journal of Social Psychology*, *43*(5), 423–437. https://doi.org/10.1002/ejsp.1958

Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and method of research*. Tavistock.

Marková, I. (2016). *The dialogical mind*. Cambridge University Press.

Mathis, A. E. (2006). General semantics and Holocaust denial. *ETC: A Review of General Semantics*, *63*(1), 52–61.

McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry*, *24*(2), 184–197.

McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of forewarning in developing resistance to persuasion. *The Public Opinion Quarterly*, *26*(1), 24–34.

McKee, M. (2006). A European alcohol strategy. *British Medical Journal*, *333*(7574), 871–872. https://doi.org/10.1136/bmj.39003.629606.BE

McKee, M., & Diethelm, P. (2010). How the growth of denialism undermines public health. *British Medical Journal*, *341*, c6950. https://doi.org/10.1136/bmj.c6950

Mead, G. H. (1934). *Mind, self & society from the standpoint of a social behaviorist*. University of Chicago Press.

Moore, H., Jasper, C., & Gillespie, A. (2011). Moving between frames: The basis of the stable and dialogical self. *Culture & Psychology*, *17*(4), 510–519. https://doi.org/10.1177/1354067X11418542

Morand, D. (2014). Using politeness to model the psychosocial dynamics of power in organizational interaction. *International Journal of Organizational Analysis*, *22*(2), 247–273.

Morant, N., & Edwards, E. (2011). Police responses to diversity: A social representational study of rural British policing in a changing representational context. *Journal of Community & Applied Social Psychology*, *21*(4), 281–296. https://doi.org/10.1002/casp.1074

Moscovici, S. (1976). *Psychoanalysis: Its image and its public*. Polity Press.

Napolitano, A. (2018). Image repair or self-destruction? A genre and corpus-assisted discourse analysis of restaurants' responses to online complaints. *Critical Approaches to Discourse Analysis Across Disciplines*, *10*(1), 135–153.

Noort, M. C., Reader, T. W., & Gillespie, A. (2019). Speaking up to prevent harm: A systematic review of the safety voice literature. *Safety Science*, *117*, 375–387.

Oborne, P., & Roberts, T. (2017). *How Trump thinks: His tweets and the birth of a new political language*. Head of Zeus.

O'Doherty, K., & Lecouteur, A. (2007). "Asylum seekers," "boat people" and "illegal immigrants": Social categorisation in the media. *Australian Journal of Psychology*, *59*(1), 1–12. https://doi.org/10.1080/00049530600941685

O'Dwyer, E., Lyons, E., & Cohrs, J. C. (2016). How Irish citizens negotiate foreign policy: A social representations approach to neutrality. *Political Psychology*, *37*(2), 165–181. https://doi.org/10.1111/pops.12242

Panagiotou, E., & Kadianaki, I. (2019). From cognitive dissonance to cognitive polyphasia: A sociocultural approach to understanding meat-paradox. *Journal for the Theory of Social Behaviour*, *49*(2), 235–253. https://doi.org/10.1111/jtsb.12201

Papastamou, S. (1986). Psychologization and processes of minority and majority influence. *European Journal of Social Psychology*, *16*(2), 165–180. https://doi.org/10.1002/ejsp.2420160205

Parker, D., Lawrie, M., & Hudson, P. (2006). A framework for understanding the development of organisational safety culture. *Safety Science*, *44*(6), 551–562.

Piaget, J. (1932). *The moral judgment of the child*. Routledge.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. Viking.

Proctor, R. N. (1995). *Cancer wars: How politics shapes what we know and don't know about cancer*. Basic Books.

Psaltis, C. (2016). Collective memory, social representations of intercommunal relations, and conflict transformation in divided Cyprus. *Peace and Conflict: Journal of Peace Psychology*, *22*(1), 19–27.

Psaltis, C., & Duveen, G. (2007). Conservation and conversation types: Forms of recognition and cognitive development. *British Journal of Developmental Psychology*, *25*(1), 79–102.

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). *Echo chambers on Facebook* (SSRN Scholarly Paper ID 2795110). Social Science Research Network. https://papers.ssrn.com/abstract=2795110

Raudsepp, M., & Wagner, W. (2011). The essentially other: Representational processes that divide groups. In I. Markova & A. Gillespie (Eds.), *Trust and conflict: Representation, culture and dialogue* (pp. 105–122). Routledge.

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2007). Interdisciplinary communication in the intensive care unit. *British Journal of Anaesthesia*, *98*(3), 347–352. https://doi.org/10.1093/bja/ael372

Sammut, G., Clark, M., & Kissaun, G. D. (2013). Dialogue, linguistic hinges and semantic barriers: Social psychological uses and functions of a vulgar term. *Journal for the Theory of Social Behaviour*, *44*(3), 326–346. https://doi.org/10.1111/jtsb.12045

Sammut, G., & Gaskell, G. (2010). Points of view, social positioning and intercultural relations. *Journal for the Theory of Social Behaviour*, *40*(1), 47–64. https://doi.org/10.1111/j.1468-5914.2009.00422.x

Sammut, G., & Sartawi, M. (2012). Perspective-taking and the attribution of ignorance. *Journal for the Theory of Social Behaviour*, *42*(2), 181–200. https://doi.org/10.1111/j.1468-5914.2011.00485.x

Schmid, P., & Betsch, C. (2019). Effective strategies for rebutting science denialism in public discussions. *Nature Human Behaviour*, *3*, 931–939. https://doi.org/10.1038/s41562-019-0632-4

Somayaji, A., Hofmeyr, S., & Forrest, S. (1997). *Principles of a computer immune system*. In Proceedings of the New Security Paradigms Workshop (Vol. 97, pp. 75–82). Association for Computing Machinery.

Stich, A., & Wagner, T. (2012). Fooling yourself: The role of internal defense mechanisms in unsustainable consumption

behavior. *North American Advances in Consumer Research*, *40*, 408–416.

Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, *37*(3), 228–243.

Sweeny, K., & Shepperd, J. A. (2007). Being the best bearer of bad tidings. *Review of General Psychology*, *11*(3), 235–257.

Tholander, M. (2011). Mundane whistleblowing: Social drama in assessment talk. *Discourse Studies*, *13*(1), 69–92. https://doi.org/10.1177/1461445610387737

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691.

Tsirogianni, S., & Andreouli, E. (2011). Beyond social cohesion: The role of "Fusion of Horizons" in inter-group solidarities. *Papers on Social Representations*, *20*, 5–1.

Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, *18*(8), 689–690. https://doi.org/10.1111/j.1467-9280.2007.01961.x

Vuori, T. O., & Huy, Q. N. (2016). Distributed attention and shared emotions in the innovation process: How Nokia lost the Smartphone battle. *Administrative Science Quarterly*, *61*(1), 9–51. https://doi.org/10.1177/0001839215606951

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049–1062.

Weick, K. E. (2010). Reflections on enacted sensemaking in the Bhopal disaster. *Journal of Management Studies*, *47*(3), 537–550. https://doi.org/10.1111/j.1467-6486.2010.00900.x

Weick, K. E., & Sutcliffe, K. M. (2003). Hospitals as cultures of entrapment: A re-analysis of the Bristol Royal Infirmary. *California Management Review*, *45*(2), 73–84.

Westrum, R. (2004). A typology of organisational cultures. *BMJ Quality & Safety*, *13*(Suppl. 2), ii22–ii27.

White, P. R. R. (2003). Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text-Interdisciplinary Journal for the Study of Discourse*, *23*(2), 259–284.

Yang, K.-C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1*(1), 48–61.

Zittoun, T., Gillespie, A., Cornish, F., & Psaltis, C. (2007). The metaphor of the triangle in theories of human development. *Human Development*, *50*(4), 208–229.