Title: Beyond Labeled Lines: A population coding account of the Thermal Grill Illusion

Authors: Francesca Fardo\textsuperscript{1*}, Brianna Beck\textsuperscript{2}, Micah Allen\textsuperscript{3,4,5}, and Nanna Brix Finnerup\textsuperscript{1}

\textsuperscript{1} Danish Pain Research Centre, Department of Clinical Medicine, Aarhus University, 8000 Aarhus, Denmark
\textsuperscript{2} School of Psychology, University of Kent, Canterbury, United Kingdom
\textsuperscript{3} Aarhus Institute of Advanced Studies, Aarhus University, 8000 Aarhus, Denmark
\textsuperscript{4} Centre of Functionally Integrative Neuroscience, Aarhus University, 8000 Aarhus, Denmark
\textsuperscript{5} Cambridge Psychiatry, University of Cambridge, Cambridge, United Kingdom

*Corresponding author

Francesca Fardo, PhD
Aarhus University Hospital
Neurologisk Forskning
J17-Plan 1, Krydspunkt J104
Palle Juul-Jensens Boulevard 165
8200 Aarhus N Denmark

Phone: +45 71790914
Email: francesca@clin.au.dk
Abstract
Heat and pain illusions can be generated by simultaneous cold and warm stimulation on the skin, at temperatures that would normally be perceived as innocuous in isolation (e.g., ‘synthetic heat’ and the ‘thermal grill illusion’). Historically, two key questions have dominated the literature: Which specific pathway conveys the illusory perception of heat and pain? Where specifically does the illusory pain originate in the central nervous system? Two major theories – the addition and the disinhibition theories – have suggested distinct pathways, as well as specific spinal or supraspinal mechanisms. However, both theories fail to fully explain experimental findings on illusory heat and pain phenomena. We suggest that the disagreement between previous theories and experimental evidence can be solved by abandoning the assumption of one-to-one relations between pathways and perceived qualities. We argue that a population coding framework, based on distributed activity across non-nociceptive and nociceptive pathways, offers a more powerful explanation of illusory heat and pain. This framework offers new hypotheses regarding the neural mechanisms underlying temperature and pain perception.

Keywords: Thermal Grill, TGI, illusion, synthetic heat, thermosensation, pain
1. Introduction
Alternating innocuous cold and warm stimuli on the skin produces unique and paradoxical sensations of strong heat and pain, often referred to as synthetic heat or the thermal grill illusion. Although each cold and warm stimulus is felt as painless in isolation, their spatial combination is often perceived as hot, burning and ‘prickling’. This peculiar sensation of illusory pain resembles the uncomfortable burning feeling that one might experience when washing cold hands under lukewarm water. This ability of simultaneous cold and warm stimuli to elicit a qualitatively new sensation, rather than two distinct sensations, or a simple sensation of their average, is at first glance counter-intuitive; why does the nervous system produce this curious quirk of sensory integration? In this article, we argue that this so-called illusion actually exposes fundamental functional principles of thermosensation, and propose a population coding framework for understanding how such illusions may illuminate the neural mechanisms underpinning temperature and pain perception in humans.

First, we review previous theories that have attempted to explain synthetic heat and the thermal grill illusion - two idioms indicating the same phenomenon, but with different emphases on the illusory temperature and pain features, respectively. The term synthetic heat denotes the dominant temperature quality and is often used in studies investigating illusory heat without pain from mild cold and warm stimulation. Synthetic heat is also known as the non-painful or heat-based thermal grill illusion (Burnett and Dallenbach, 1927; Green, 2002, 1977). In contrast, the thermal grill illusion (TGI) is named for the type of stimulation apparatus used to elicit the illusory heat and pain sensations - a grill of alternating cold and warm bars, which, when placed on the skin, creates a thermnociceptive prickling sensation. Thermal grills were originally developed to investigate synthetic heat (Bishop, 1927; Cutolo, 1918; Twitmyer and Fernberger, 1927), before becoming the conventional method to investigate illusory pain (e.g., Craig and Bushnell 1994; Craig et al. 1996; Bouhassira et al. 2005; Leung et al. 2005). We henceforth use the term thermal grill illusion (TGI) in the more general sense to describe both painless and painful heat sensations elicited by the spatial alternation
of cold and warmth. Instead, we use the terms illusory heat or illusory pain when referring to either the thermal or the painful aspects of the TGI, specifically.

What are the neurophysiological mechanisms underpinning these phenomena? Two major theories have dominated the debate in the past 25 years: the addition or convergence theory (Bouhassira et al., 2005; Green, 2002) and the thermosensory disinhibition or unmasking theory (Craig and Bushnell 1994; Craig et al. 1996; Craig 1998). In this opinion article, we first provide a critical review of the evidence in support of these theoretical perspectives. Critically, these classical theories are built upon the inter-related notions of sensory specificity and labeled line coding in the thermosensory and nociceptive systems. We thus begin by reviewing the extant evidence for each of these core principals, before arguing that a population coding account of temperature and pain perception can more easily account for TGI and related phenomena.

2. TGI phenomena
The paradox of the thermal grill illusion lies in the mismatch between the applied sensory inputs and the seemingly unrelated perceptual outcome. In the case of illusory pain, this mismatch is between the non-noxious thermosensory inputs and the subjective perception of pain. This has puzzled generations of scientists since the early observations by Thunberg (1896) and Alrutz (1898), who independently discovered the phenomenon. Following the discovery of the TGI, the basic existence of the illusion was questioned (Jenkins, 1938a, 1938b), leading to a long period of research quiescence. Interest in TGI was later revived by the influential work of Green (Green, 2002, 2004, 1977), Craig (Craig, 1998; Craig and Bushnell, 1994), Bouhassira (Bouhassira et al., 2005a), and Defrin (Defrin et al., 2008), eliminating any doubt as to whether illusory heat and pain are reproducible phenomena.

To induce TGI sensations, many studies use thermal grill devices, with a variable number of alternating bars which can be individually cooled or heated to produce uniform or alternating patterns
of thermal stimulation (Bouhassira et al., 2005a; Craig and Bushnell, 1994; Harper and Hollins, 2017, 2014; Leung et al., 2005). TGI stimuli produce subjective sensations predominantly characterized by thermal sensations in the warm-to-hot range. When pain is experienced, its intensity is typically mild to moderate (Bouhassira et al., 2005a; Craig and Bushnell, 1994; Leung et al., 2005). However, similar heat and painful percepts can be elicited using different experimental apparatuses, such as alternating cold and warm spiral tubes (Bach et al., 2011; Thunberg, 1896), warm-cold-warm patterns using three thermodes (Kammers et al., 2010; Marotta et al., 2015), or one single cold and one single warm thermode, even at large distances on the skin (Defrin et al., 2008; Fardo et al., 2018). Overall these experiments suggest that the TGI depends primarily upon the spatial pattern of thermal stimulation, while other factors such as the size of the affected area, the number of stimulated afferents, or the distances between component stimuli exert minor influences on the quality and intensity of the illusion (Li et al., 2009).

A key parameter to induce TGI is the degree of temperature difference between the simultaneous cold and warm temperatures (Burnett and Dallenbach, 1928, 1927; Gritman and Dallenbach, 1929). Smaller differences (e.g., 10-15 °C) are sufficient for illusory heat, while larger differences (e.g., ≥20 °C) are required for illusory pain sensations (Adam et al., 2014; Bouhassira et al., 2005a). Thus, although the generation of TGI percepts is conserved across a variety of spatial configurations, the difference between the cold and warm components does exert a large influence on the perceived thermo-nociceptive quality and intensity of the illusion.

TGI phenomena have been investigated using qualitative reports - for example by requiring participants to select from a variety of descriptors relating to the subjective qualities of temperature and pain (e.g., Li et al. 2009; Bach et al. 2011) - as well as quantitative approaches, such as verbal numerical scales (Scheuren et al., 2014), visual analogue scales (Adam et al., 2014; Craig and Bushnell, 1994; Kern et al., 2008a, 2008b; Leung et al., 2005) or temperature matching (Fardo et al.,
2018; Ferrè et al., 2018; Kammers et al., 2010; Leung et al., 2005; Marotta et al., 2015). From this variety of qualitative and quantitative measures, it emerged that TGI sensations vary from innocuous to painful heat percepts, and can be felt as an intermediate sensation between innocuous and noxious perception (Bach et al., 2011). The interpretation that TGI sensation is a gradient between thermally neutral and thermally painful percepts is echoed by Craig, who wrote that specific neurons involved in TGI perception “do not merely signal ‘pain’ and must have another role”, and further suggested that this role could be “a gradually increasing feeling of discomfort that motivates an appropriate thermoregulatory behavioral response” (Craig 2014, pg. 74). While there is emerging agreement on the functional interpretation of TGI as a gradient, or possibly an independent dimension from nociceptive pain (Alfonsi et al., 2016), it remains a topic of intense debate as to which neurophysiological mechanisms underpin illusory pain.

Previous theories largely focused on which specific unmyelinated fibers drive the perception of illusory pain (Bouhassira et al., 2005; Craig and Bushnell, 1994; Green, 2002). Both the main theories assume that the illusion is driven by C-fibers, because of the slow and delayed onset of the illusion (a few seconds) and the sustained response over time (Hunter et al., 2015), which are compatible with the temporal delays and sustained activity of C-fibers. Further, the frequency of warm and hot descriptors associated with thermal grill stimuli was found to be unaffected by the blockage of thinly myelinated Aδ-fibers, suggesting that C-fibers are the main contributors (Fruhstorfer et al., 2003). However, the discrepancy between the major theories is related to which specific class of C-fiber is involved: wide dynamic range (WDR) neurons (Bouhassira et al., 2005; Green, 2002) or C-polymodal nociceptors (Craig, 1998; Craig and Bushnell, 1994).

A related topic of debate is how and where cross-talk between thermosensory and nociceptive inputs occur in the nervous system. In other words, where in the central nervous system do these afferent pathways interact to yield the “painfulness” of the TGI. So far, the prevailing assumption is
that thermo-nociceptive interactions must occur centrally on account of the observation that mild warm and cold peripheral inputs are by definition non-nociceptive. As such, the debate has focused on the distinction between supraspinal (thalamus or cortex; Craig and Bushnell 1994) and spinal mechanisms (Bouhassira et al., 2005; Green, 2002). However, experimental data do not support this categorical approach, and indicate that we cannot exclude either spinal (Fardo et al., 2018; Harper and Hollins, 2017) or supraspinal contributions (Ferrè et al., 2018; Lindstedt et al., 2011; Marotta et al., 2015) as contributing to the TGI phenomenon. In the next paragraphs, we will review the two main TGI theories and their assumptions in more detail. We will then outline alternative assumptions and propose a new population coding theory, which can better account for the neuroanatomical, neurophysiological and psychophysical data at hand.

3. Neurophysiological theories of the thermal grill illusion

The two major TGI hypotheses - the addition and disinhibition theories - are both influenced by the specificity theory of somatosensation, which has its roots in the law of specific nerve energies as developed by Müller in 1826 (for a review on the history of labelled lines, see Norrsell et al. 1999). This law postulates that each sensory quality corresponds to a distinct afferent pathway, associated with specific receptors in the periphery and cortical targets in the brain, which determine the quality of perception. The specificity theory thus suggests that somatosensation is coded by labeled lines, namely, pathways dedicated to each skin sense (e.g., cold, heat, pain and touch). For example, the theory assumes that the nociceptive pathway is uniquely responsible for pain perception and is activated by noxious stimulation of specific nociceptors transmitting a signal from the periphery to specific spinal and supraspinal projection neurons (for a review, see Perl 2007), as if following a ‘labeled line’. On this view, the conundrum that previous TGI theories aimed to explain was how
cold and warm non-nociceptive peripheral afferents can lead to percepts with distinct heat and nociceptive qualities, assumed to be associated with a different labeled line.

3.1 The addition theory

The addition or convergence theory (Bouhassira et al., 2005; Green, 2002) suggests that illusory pain is produced by a polymodal pathway that encodes the linear summation of warm and cold sensory inputs (Fig. 1A). Originally, the additive property of cold and warm integration was derived from psychophysical experiments in humans. These results demonstrated that increments in synthetic heat perception depend upon the temperature differential between cold and warm temperatures (Burnett and Dallenbach, 1927). The strength of the synthetic heat on the continuum from warmth to heat and pain was thought to be primarily determined by the absolute difference between cold and warm sensory inputs (i.e., the greater the temperature difference, the greater the neural summation). To account for the prevailing heat quality, previous authors proposed mathematical models based on a simple additive function, where the warm addend is weighted 3 times more than the cold addend (Burnett and Dallenbach, 1928; Gritman and Dallenbach, 1929). Under this view, both warm and cold temperatures activate the same polymodal pathway, leading to summation of individual inputs irrespective of their cold or warm temperatures. The increased firing rate, due to summation, is assumed to mimic the firing rate associated with more intense stimulation typically perceived as painful. Thus, the strength of the resulting sensation would be directly related to the strength of the combined cold and warm afferent drive.

Proponents of the addition theory suggested that the putative polymodal pathway might originate from second-order wide-dynamic range (WDR) neurons in the deep layers of the dorsal horn in the spinal cord (Bouhassira et al., 2005; Green, 2002). Spinal WDR neurons respond to Aβ, Aδ and C-fiber peripheral inputs (Mendell, 1966) generated by mechanical and thermal stimulation
Fardo et al.  
Population coding of TGI

(Khasabov et al., 2001). The potential involvement of WDR neurons in producing the painful quality of the thermal grill illusion is consistent with the role of WDR neurons in sensory and discriminative processing of noxious stimulation (Coghill et al., 1993; Maixner et al., 1986), as well as sensitization phenomena associated with clinical pain conditions (Hao et al., 1992; Schouenborg and Sjolund, 1983; Svendsen et al., 1997). On this view, the recruitment of WDR neurons accounts for the painful quality of the thermal grill illusion, in a similar fashion as it accounts for the painful quality of properly noxious stimulation.

3.2 The disinhibition theory

To explain illusory TGI pain, the disinhibition or unmasking theory (Craig, 1998, 2003; Craig and Bushnell, 1994) posits certain interactions between labeled lines. Specifically, it assumes two distinct serial steps. First, warm stimulation inhibits the response of COOL neurons in the spinal cord (i.e., neurons responding to mild cold temperatures <27 °C), precisely in the superficial laminae of the dorsal horn. Second, C-polymodal nociceptors (i.e., heat-pincha-cold or HPC neurons) are disinhibited at the thalamo-cortical level, due to reduced ascending inhibition from COOL neurons. In other words, the spinal “warm inhibits cold” mechanism lessens the thalamo-cortical “cold inhibits pain” mechanism, leading to “disinhibition of pain” (Fig. 1B).

The “warm inhibits cold” assumption was based on electrophysiological recordings of spinal dorsal horn neurons in anesthetized cats. Craig found that when a 20°C cold stimulus was delivered in isolation, both COOL and HPC neurons were active, but COOL neurons responded more strongly than HPC neurons. However, if both 20°C and 40°C were simultaneously applied, the activity of COOL neurons was reduced by 50%, while HPC activity was unaffected (Craig and Bushnell, 1994). The authors proposed a simple physiological model, where thermal grill percepts depend upon the difference between HPC and COOL neuron activity (HPC − COOL). A negative difference (COOL
> HPC) corresponds to sensations of innocuous cold, while a positive difference (HPC > COOL) corresponds to sensations of burning pain.

Compared to the “warm inhibits cold” mechanism, the assumption that “cold inhibits pain” at the thalamo-cortical level was based on observations that ischemic and compression nerve blocks differentially alters cold, warm and pain perception (Fruhstorfer, 1984; LaMotte and Thalhammer, 1982; Mackenzie et al., 1975; Wahren et al., 1989; Yarnitsky and Ochoa, 1990). Specifically, these studies suggested that the perception of cold, warmth and pain might be mediated by fibers with different amounts of myelination. Innocuous cold is thought to be primarily mediated by thinly myelinated Aδ-fibers, since cold but not warm sensations cease after selective Aδ block. Instead, warm and burning pain sensations are thought to be mediated by C-fibers, given that these sensations are abolished by chemical block of unmyelinated fibers (Fruhstorfer et al., 2003). Since Aδ block enables burning pain sensations to be perceived at higher cool temperatures (e.g., 25°C rather than 15°C), previous authors have inferred that Aδ-fibers (nominally “cold fibers”) ought to exert an inhibitory influence on C-fibers (nominally “pain fibers”) under normal circumstances.

Finally, the assumption that cold inhibits pain at a thalamo-cortical level was justified by the observation that spinal HPC activity was unaffected by simultaneous cold and warm stimulation in cats (Craig and Bushnell, 1994). This assumption was considered compatible with evidence that the thalamus contributes to the generation of pain sensations, given that lesions of its posterior part can cause phantom sensations of burning pain in a contralateral skin area with reduced sensitivity to thermal and nociceptive stimulation (Head and Holmes, 1911).

Although the additive and disinhibition theories provide different neurophysiological assumptions, both posit that the recruitment of a specific population of neurons (i.e., WDR or HPC) accounts for the perception of pain irrespective of whether the peripheral input is nociceptive or not.
We henceforth refer to these theories collectively as specificity theories in virtue of their specificity assumption (i.e., one pathway = one percept), as opposed to population coding theories.

4.3 A population coding framework

An alternative hypothesis to the previously described theories can be derived from population coding models, originally developed in the field of vision neuroscience (Georgopoulos et al., 1986). This proposal is reminiscent of the pattern theory of color vision put forward by Young (1802) and von Helmholtz (1860), and the across-fiber pattern approach discussed by Erikson (Erickson, 1973, 1968; Erickson and Poulos, 1973). The key concept of population encoding is that perception is driven probabilistically by the joint response of populations of neurons (for reviews, see (Averbeck et al., 2006; Panzeri et al., 2015; Quiroga and Panzeri, 2009). Each individual neuron contributes to a different extent, depending on the strength of its response relative to the rest of the population. On this view, the nervous system integrates inputs from populations of neurons to provide a summarized output, based on the distribution of the neuronal responses. Key parameters could be the mean and variance of population activity, as determined by the individual stimulus-response functions derived from each subset of neurons with similar response properties. Population encoding proposes that some sensory feature – such as the quality or intensity of a thermal stimulus – is encoded by the mean activity of a range of populations of neurons, weighted by the variability within each population. A fascinating and powerful feature of this mechanism is that the encoded perceptual features depend upon the probabilistic activity of a population of neurons, and that stimulus uncertainty (i.e., signal-to-noise ratio) can be naturally accounted for by the variance across neuronal populations. Given that distributed encoding schemes do not depend solely upon individual neurons or a unique population of neurons, but on the distribution of responses across several populations of neurons, this type of
encoding safeguards the robustness of the computed outputs against uncertainty and neuronal variability.

Here, we propose a model of thermo-nociceptive encoding where each processing stage and neuronal population is characterized by distinct population codes (i.e., stimulus-response functions) whose multiplex output encodes the probability of potential tissue damage across thermosensory modalities. On this view, temperature and pain perception depend on spatio-temporal integration of the responses of several classes of thermo-nociceptive neurons at several processing levels (Fig. 2). For simplicity, we depict three major processing stages (peripheral, spinal and supraspinal) and three neuronal populations (warm-responding vs. cold-responding vs. polymodal neurons). We assume integration at each processing level. In the periphery, some degree of integration can be achieved via ephaptic coupling among unmyelinated fibers. In the spinal cord, integration can occur between mechanosensory, thermosensory and nociceptive modalities via interneurons. At higher processing stages, multisensory integration can occur between distinct senses (e.g., vision and nociception) via short-range and long-range anatomical and functional connectivity across brain networks.

On the basis of this model, we propose that the degree to which we experience innocuous and noxious thermal percepts is represented by a unique distribution of neural activity across thermosensory and nociceptive neurons in a distributed fashion throughout the nervous system. From the perspective of thermo-nociceptive population encoding, we hypothesize that the TGI is not an anomalous misperception of combined cold and warm stimulation, but instead as the normal consequence of a particular pattern of neuronal activity which the brain interprets as a sensation of heat or burning pain.

It is crucial to note that the outlined architecture is a simplistic sketch to illustrate some key principles of the proposal, and we intentionally refrain from making predictions on the specific putative response functions describing the “population code” for different type of percepts. This is
because the aim of this paper is to generate discussion and provide a framework for motivating new research questions and hypothesis on the TGI and possibly other thermo-nociceptive phenomena. Under this framework, the questions “which specific pathway does convey the illusory perception of heat and pain? And where specifically does the illusory pain originate in the central nervous system?” are ill-posed questions. A more promising approach would be to ask which distributed patterns underlie TGI percepts and how such patterns are differentiated from those underlying innocuous thermosensation or pain perception. Other key questions are related to which neurobiological mechanisms (e.g., lateral inhibition or other spatial and temporal-contrast enhancing mechanisms) are at place to enable the perception of the rich repertoire of thermal and pain percepts that we normally experience. From the perspective of information theory, one can then ask: what are the critical modes of information, the associated transformations governing their integration (e.g., linear or nonlinear), and which specific features are integrated and communicated at distinct processing stages through networks of ascending and descending neurons? The emphasis here is not on the one locus in the nervous system that is responsible for a specific percept, but rather on how the nervous system is organized and which type of information is extracted and propagated at different processing levels to ultimately construct thermal and pain perception.

4.4 A population coding example

An exemplary approach can be found in a recent article from Wang and collaborators (2018), which demonstrated different coding properties of dorsal root ganglion (DRG) neurons (i.e. first-order neurons) responding to cold or warm stimuli in mice. They demonstrated that the response profile of heat-responding DRG neurons is compatible with graded population coding, as neural activity increased with temperature increments both at the single-neuron (i.e., how strongly a neuron responds) and population levels (i.e., how many neurons are activated). This is a simple form of
population coding, compatible with vector coding, previously described in the visual cortex (Georgopoulos et al., 1986). The critical parameter is the mean neuronal activity, which can be conceptualized as the sum of the responses of a population of neurons, weighted by the firing rate of each individual neuron in the population. This view assumes rate coding at the level of individual neurons (i.e., the firing rate of each neuron encodes one value) and vector coding at the level of the population (i.e., the mean of the population neurons encode a summary value with high signal-to-noise ratio). In contrast, the response profile of cold-responding DRG neurons showed a striking difference to heat-responsive DRG neurons. Cold-responding neurons activity was better fit by combinatorial population coding, due to ungraded neural activity with increased cold temperatures (i.e., lack of linear increase of neural activity with increased cold temperatures) (Wang et al., 2018).

The observations that peak amplitude, decay and rise time of cold-sensitive neuron responses do not reflect the intensity of cold temperature indicates that the propagation of cold-related inputs to the spinal cord and, possibly, the further interpretation of such signals as cold perception depends on the pattern of co-activated afferent neurons. Although these results clearly illustrate the idiosyncratic properties of heat and cold-responsive DRG neurons in mice, it is still unknown how the responses of heat-, cold- and dually-responsive neurons are integrated in the central nervous system in the case of combined warm and cold stimulation as in the TGI paradigm, and whether this population coding schemes are preserved across species.

5. Specificity vs. population coding

The long-standing puzzle of the thermal grill illusion can be explained by the inability to reconcile the assumptions of specificity theories, postulating physiologically and perceptually independent labeled lines, with state-of-the-art knowledge on neuroanatomy, neurophysiology and perception in humans and animals. In the next section, we outline how specificity theories fail to account for many
neuroanatomical and functional properties of the mammalian thermo-nociceptive system while outlining how our population coding model can better account for these shortcomings.

5.1. Neuroanatomical argument
One key assumption of specificity approaches, such as the addition and the disinhibition theories, are the existence of specific and independent pathways connecting the periphery to the brain. According to the specificity principle, each pathway or labeled line directly relays neural activity from peripheral receptors to the cortex. Neural activity along a labelled line can be blocked or facilitated (e.g., cross-talk), but the functional meaning of the labeled line is faithfully preserved throughout the nervous system.

If this view was correct, one would assume the existence of first-order neurons expressing one specific type of transducers, forming one-to-one synapses with other specific neurons in the central nervous system. However, not all first-order neurons are selectively tuned to encode one specific perceptual category. In contrast, a large extent of first-order neurons are polymodal, as they co-express different transducers sensitive to different thermal and/or mechanical changes (e.g., McKemy et al. 2002). In the central nervous system, distinct neuronal populations largely converge and interact within spinal microcircuits (for reviews, see; Todd 2010; Prescott and Ratté 2012; Benarroch 2016). Only a portion of spinal neurons travel to the thalamus (i.e., projection neurons), while other neurons are confined within the spinal cord (i.e., propriospinal neurons) (Kerr, 1975). The subset of projection neurons are thus most likely encoding a new output that reflects integration between several population of neurons at the spinal level, rather than relaying information that is isomorphic to the sensory input or the perceptual outcome. Further, with regard to the neuroanatomical specificity of pain, experimental data from brain imaging and intracranial recordings clearly show that there is no positive evidence of single areas mediating pain perception in the brain (for a review, see Mouraux and Iannetti 2018). This organization suggests that neural activity is not
faithfully relayed from the periphery to the brain but is instead substantially transformed at spinal and supraspinal processing stages.

To summarize, the heterogeneous co-expression of distinct transducers in first-order neurons, the large degree of convergence on spinal and supraspinal neurons, and the large extent of lateral interactions in propriospinal and supraspinal circuits are all incompatible with a view postulating direct and specific pathways from receptors to cortex. Particularly for pain, there is also a lack of evidence for any brain area specific to its processing. While specificity theories struggle to explain these neuroanatomical properties, a population coding model can easily account for them, given that the overall encoded sensory feature is a probabilistic aggregate of multiplex stimulus-response profiles of neuronal populations at different processing stages.

5.2 Neurophysiological argument

The addition and the disinhibition theories assume a one-to-one correspondence between specific pathways and the perception of illusory heat and pain. However, illusory sensations can be elicited by stimuli that are outside of the physiological ranges of the specific neurons assumed to mediate the illusion (i.e., WDR and HPC neurons, respectively). Craig and collaborators suggested that WDR neurons could not explain the TGI, given that WDR neurons respond to cold only at temperatures below 15°C (Craig and Bushnell, 1994). Using a similar argument, other authors considered the disinhibition theory unlikely (Defrin et al., 2008; Fruhstorfer et al., 2003; Green, 2002), given that HPC neurons only respond to cold at temperatures below 19-22°C in humans (Campero et al., 1996). Both WDR and HPC neurons are thus silent at temperatures that are sufficient to elicit illusory heat (Green, 2002). Further, the strength of the TGI can be modulated by integration with other sensory modalities, and by manipulation of the spatial configuration of body parts (Kammers et al., 2010; Marotta et al., 2015; Seckel et al., 2012). The lack of a clear one-to-one relationship between a
specific fiber type and illusory heat and pain, as well as transformation of thermal and pain percepts through multisensory integration, cannot be easily reconciled with specificity theories. On the other hand, population coding does not rely on one specific fiber type, but rather on the joint activity across populations of neurons. On this view, thermo-nociceptive as well as multisensory integration correspond to the extrapolation of refined and noise-robust output signals at various levels of the neuroaxis. These processes can be explained by convergence and lateral interaction mechanisms.

5.3. Perceptual argument

From a perceptual point of view, neither the additive nor the disinhibition theories can account for the peculiar thermo-nociceptive qualities of illusory pain. If TGI was truly driven by WDR or HPC neurons in a labeled line fashion, the quality of the experience should be indistinguishable from the common experience of thermal pain. Instead, the TGI is normally experienced as a unique blend of thermal and nociceptive qualities that is not fully identical to either hot or cold pain. The sensation is better described as a gradient between neutral and noxious perception (Bach et al., 2011).

The classic one-to-one assumption between sensory inputs and perceptual outcomes can be criticized based on the observation that perception does not only depend upon the activity within single channels. For instance, increasing warm and heat sensitivity via capsaicin (i.e., targeting heat-responsive TRPV1 receptors) also decreases cold sensitivity (Callsen et al., 2008). Similarly, increasing cold sensitivity via menthol (i.e., targeting cold-responsive TRPM8 receptors) also decreases warm sensitivity (Green, 1986). Further, the selective impairment of cold and warm innocuous thermal sensations was shown to affect the perceived quality and perceptual threshold of nociceptive events in spinal cord injured patients (Defrin et al., 2002). These findings are often explained by linear inhibition models that assume “warm inhibits cold” or “cold inhibits pain” type of interactions (e.g., Craig and Bushnell 1994). In these models, the interaction between cold and
warm pathways proceeds solely via inhibition or disinhibition, while the perceptual outcome still depends on one specific pathway. Although this explanation might sound intuitive, it does not explain how the activation of the same single pathway can generate a varied range of perceptual outputs, which can be at the intersection between distinct qualities.

One possible objection here is that our criticism of the supposed one-to-one relationship between sensory inputs and perceptual outcomes is incompatible with microneurography results using intraneural microstimulation (INMS) of a single afferent fiber. In some cases, INMS elicit a spatially localized sensation whose quality depends upon the type of fiber stimulated (e.g. flutter for RA1, or burning for a C-nociceptor; Ochoa 2010). This suggests that activation of a single peripheral somatosensory neuron might be sufficient to elicit a conscious sensation. However, we rebut this objection to thermo-nociceptive population coding on several grounds. First, it is important to remember that population coding can occur at multiple levels of the neuroaxis. Even if a single first-order neuron can be selectively stimulated, that neuron synapses with multiple second-order neurons in the spinal cord, and those neurons, in turn, may activate multiple neurons supraspinally. Indeed, INMS of a single mechanoreceptive afferent has been shown to activate whole networks of somatosensory and cortical areas (Sanchez Panchuelo et al., 2016; Trulsson et al., 2001). Thus, a population code is likely still operating in the central nervous system during peripheral INMS. Second, some features of normal sensation are difficult to account for by activation of a single peripheral afferent alone. INMS tends to elicit ‘quantal’ sensations, meaning that the perceived intensity of the sensation does not increase in a graded fashion with increments in stimulus energy. Thus, perception of sensory intensity may not be easily accounted for by the activity of an individual first-order sensory neuron. Moreover, INMS has so far only elicited very simple, ‘pure’ sensations such as tapping or vibration. We propose that naturalistic, complex sensations - and even hybrid
sensations like the TGI - depend upon the patterns of co-activation across populations of sensory neurons.

To conclude, we argue that temperature and pain perception in general - and, more specifically, perceptual phenomena such as the thermal grill illusion - are more easily explained by adopting a population coding perspective. Under this view, the thermosensory system does not have to encode each possible sensory quality and intensity using separate neurons and labeled lines. Instead, it can generate graded or unique responses by interpolation based on the collective activity of populations of neurons. This view offers a way to explain “illusions”, not as exceptional circumstances under which the system is at fault, but as an expression of a fully functioning sensory system.

6. Conclusions
Population coding has been previously hypothesized to explain perception of spatial and intensity properties of nociceptive stimulation (Quevedo and Coghill, 2009), as well as more generally within the somatosensory system (Doetsch, 2000; Ma, 2012, 2010). However, so far, very few empirical studies have addressed specific population coding mechanisms in the thermo-nociceptive system (Wang et al., 2018). Here, our primary goal was to illustrate how the population coding framework can overcome the limitations of specificity theories with respect to their neurobiological plausibility and their potential to account for experimental evidence on temperature and pain perception in humans. To illustrate the empirical potential of this approach, we also sketched out some basic hypothetical parameters of a population encoding account of TGI, as a starting point for generating new questions for empirical research. We stress, however, that this account is far from complete. Indeed, many details are still needed to deliver a comprehensive encoding model of temperature and pain perception in humans. For example, within the population coding framework, which sub-populations of neurons are necessary for the perception of specific thermal and nociceptive qualities?
What is the type of information encoded by different neurons at different processing stages (i.e., peripheral, spinal, and supraspinal)? How are networks of neurons spatially organized? How is information integration between neurons temporally organized? And what are the consequences of altering the distinctive spatial and temporal organization of neuronal networks in terms of perceptual outcomes?

The strength of the proposed population coding as a framework for the TGI is precisely this capacity to provide novel empirical hypotheses which can be tested using the tools and methods of computational and theoretical neuroscience. Computational models can be used to describe neural activity at different levels of the neuraxis, sampled with existing neuroimaging techniques or emerging tools, such as population receptive field modelling and large-scale population recording techniques (e.g., neuropixel). Further, mechanistic computational models can be developed using simulations of single neurons and networks. The parameters of these models, such as the encoding variability of warm and cold afferents, can further be mapped to individual differences in TGI perception and clinical disruptions thereof.

One intriguing interpretation arising from our model is that cold and warm spatial alternations are not an unusual pattern of stimulation that the brain does not know how to interpret. Instead, cold-warm temperature differentials are a key feature driving the subjective perception of potential harm. When one gets hurt, the temperature difference between injured and uninjured tissue may be one key feature that the nervous system samples to localize where a possible damage is most likely about to occur. In other words, the central nervous system might encode potential tissue damage not only in terms of absolute skin temperature, but also in terms of the relative difference in skin temperature between a particular skin region and adjacent regions. This approach would ultimately illuminate why humans experience the thermal grill illusion, while also shedding light on basic computational mechanisms of pain and thermosensation themselves.
Acknowledgements

Preparation of this article was supported by the Danish Council for Independent Research (6110-00643B, FF). MA is supported by a Lundbeckfonden Fellowship (R272-2017-4345), the AIAS-COFUND II fellowship programme that is supported by the Marie Skłodowska-Curie actions under the European Union’s Horizon 2020 (Grant agreement no 754513), and the Aarhus University Research Foundation.
Box 1. Main characteristic of a hierarchical population coding scheme

**Probabilistic.** Neurons are stochastic by nature. When the same neuron or a group of neurons are activated multiple times by the same stimulus, the neural responses (i.e., firing rate) are never identical. Extracting signal from noisy neurons requires by necessity a probabilistic rather than a deterministic process. Population coding schemes are robust to neural variability given that they do not faithfully rely on single neurons, but are instead assumed to estimate a probability distribution based on the pattern of neural responses across several integrated populations of neurons.

**Robust to noise.** Generally, even the simplest stimuli activate, at the very least, hundreds to thousands of neurons, and are associated with substantial neural variability. Not all active neurons are actual signal; instead, part of the variability in neural responses consists of noise. Population coding schemes enable the separation of signal from noise, depending on patterns of correlated or uncorrelated neural responses across populations of neurons.

**Spatially organized.** The spatial organization of neuronal populations contributes to the properties of population codes, and can assume distinct forms in terms of topographic mapping, as well as intralaminar and translaminar connectivity in the spinal cord and supraspinally (e.g., cortical columns). One key property is lateral inhibition between neighboring population of neurons. This offers added efficiency in terms of encoding more abstract features of sensory stimuli through center-surround modulation and similar spatial mechanisms.

**Temporally organized.** Population codes leverage not only the spatial organization of populations of neurons, but also the temporal unfolding of neural responses. Tactile, thermal and nociceptive qualities are typically associated with distinct fiber latencies (i.e., temporal delays related to when a signal reaches the cortex). It is reasonable to assume that distinct latencies are intrinsic properties of population codes underlying the perception of distinct somatosensory qualities.
Fig. 1 Classic specificity theories of the TGI. (A) Perception of cold stimulation and thermal grill illusion according to the addition hypothesis (Bouhassira et al., 2005; Green, 2002). Cold and warm-related sensory inputs are summated at the level of second-order neurons. The spinal response function is ‘cold plus warm’. The greater the summation, the greater the illusion of synthetic heat and pain. Indeed, the increased firing rate, due to summation, might mimic the firing rate associated with more intense stimulation typically perceived as painful. The proposed neurons that can account for such summation, as well as the painful quality, are WDR neurons. (B) Perception of cold stimulation and thermal grill illusion according to the disinhibition hypothesis. This theory is based on three interlinked assumptions: “warm inhibits cold”, “cold inhibits HPC/pain”, and thus “HPC/pain is disinhibited” in thalamus/cortex”. Specifically, the thalamus acts as a comparator between the mean discharge of HPC neurons and the mean discharge of COOL neurons (Craig and Bushnell 1994; Craig 2003). The thalamic response function is thus ‘HPC minus COOL’, leading to disinhibition of pain.
**Figure 2**

Fig. 2 Population coding hypothesis. The key concept of the population coding hypothesis is that perception depends upon neural activity across distinct populations of neurons at various levels of the neuraxis. The figure depicts the hypothesis that perception depends on spatio-temporal integration of several classes of thermo-nociceptive neurons, and that different population codes operate at distinct levels. The model is sketched as an artificial neural network, with distinct levels corresponding to first-order (i.e., Dorsal Root Ganglion), second-order (i.e., spinal) and third-order (i.e., supraspinal) neurons. For simplicity, we represent three distinct thermo-receptive populations; however, more populations are likely involved. In red, we depict warm- and heat-responding neurons; in blue, cold-responding neurons; in green; heat and cold responsive neurons (i.e., dually-tuned). At the supraspinal level, thermo-nociceptive and other sensory signals are integrated (i.e., multisensory integration). Putative response functions are indicated for each level of the neuroaxis and functionally distinct population of neurons. The probability (P) of neural responses is depicted as a function of stimulation temperature (T). Under this account, the TGI is not considered an anomalous misperception of cold stimulation, but the normal consequence of a particular pattern of neuronal activity which the brain interprets as a sensation of heat or burning pain.
References


