The Philosophy of Temperature Perception

A.J. van Westen

Supervisor **Prof. dr. Bence Nanay**

Thesis submitted for the degree of Doctor of Philosophy
Faculty of Arts | Antwerp, 2023

University of Antwerp

Universiteit Antwerpen

Faculteit Letteren en Wijsbegeerte
Departement Wijsbegeerte

The Philosophy of Temperature Perception

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Wijsbegeerte aan de Universiteit Antwerpen te verdedigen door
Allert Jaap van Westen

Promotor:
Prof. dr. Bence Nanay

Antwerpen, 2023

Jury:
Promotor: Prof. Dr. Bence Nanay (Universiteit Antwerpen)
Prof. Dr. Kathleen Akins (Simon Fraser University)
Dr. Richard Gray (University of Cardiff)
Prof. Dr. Bert Leuridan (Universiteit Antwerpen)
Prof Dr. Judith Martens (Universiteit Antwerpen)

Funded by the European Union (ERC, STYDS, 726251). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgements

Over the course of the six years that I have been working on this thesis, there have been ups and downs in the process. I can account for most of the downs myself. For the ups I need to thank some people. I consider myself extremely lucky to have people around me that supported me and helped me finish this project.

First and foremost, I need to extend my gratitude to Bence Nanay, who hired me as a PhD student on his ERC funded project *Seeing Things You Don't See*. He has been the most patient and supportive supervisor one could wish for. Sometimes, after sweating over a draft of a chapter for weeks, I would send it to Bence hoping for a few days off before getting feedback. Often, he would send back extensive and extremely insightful comments within hours. I have learned a great deal about philosophy, writing, and academia from these comments as well as from the example Bence sets as a philosopher.

Secondly, I am thankful to the members of the jury who have graciously agreed to read this thesis, assess it, and discuss it at the public defense: Kathleen Akins, Richard Gray, Bert Leuridan, and Judith Martens. I would also like to thank the members of the individual doctoral committee, Erik Myin and Peter Reynaert, for reading and evaluating this thesis.

One of the best things about being a PhD student at the Antwerp Center for Philosophical Psychology has been getting to work (and hang out) with all the members of the research group: Solveig Aasen, Nicolas Alzetta, Brandon Ashby, Alma Barner, Constant Bonard, Adam Bradley, Denis Buehler, Santiago Echeverri, Peter Fazekas, Christopher Gauker, Loraine Gérardin-Laverge, Harmen Ghijsen, Kris Goffin, Magdalini Koukou, Beate Krickel, John Kulvicki, Kevin Lande, Jason Leddington, Lucien Leigh, Francesco Marchi, Manolo Martinez, Chris McCarroll, Stephen Müller, Mirja Perez de Calleja, Nicolas Porot, Thomas Raleigh, Brad Saad, Carlota Serrahima, Laura Silva, Sam Taylor, Gerardo Viera, Dan Williams, Nick Wiltsher, Benjamin Young. Their comments on drafts and the discussions we've had have informed every chapter in this thesis.

I am also thankful to my previous philosophical mentors, John Heil and Albert Visser, who encouraged me to pursue a PhD in philosophy. I am indebted to Casey O'Callaghan for introducing me to the philosophy of perception during my stay (as part of my master's program) at Washington University in St. Louis. I am also grateful to Mohan Matthen for the opportunity to attend the International Multisensory Research Forum in Toronto in 2018.

Many thanks to Ton Haak for proofreading the thesis.

I would also like to thank my circle of philosophical friends in Utrecht: Daan Dronkers, Andries de Jong, Niels van Miltenburg, Dawa Ometto, and Clint Verdonschot. Special thanks to my long-time buddy Niels Wildschut who used to read (and critique) just about everything I wrote.

I couldn't have completed this thesis without the support and welcome distractions provided by my friends outside of philosophy. I hope to see you at the party after the defense.

The love and support from my family has been invaluable to me. I would like to thank my sister, Veerle, and my parents Thea and Jeroen for nurtring creativity and curiosity.

More than anyone else, my partner Lidewij has been by my side while writing this thesis. Even though I make the same temperature joke almost every day, she still shows me nothing but love. Thank you.

The most welcome distraction of all showed up near the end of writing this thesis: Eline, you are the best.

Table of contents

Introductie

Temperatuurperceptie als een filosofisch onderwerp

Het lijkt misschien vreemd om een proefschrift in de wijsbegeerte te schrijven over een specifiek zintuig. Zou filosofie niet moeten gaan over 'grote' vragen, en de abstracte, a priori antwoorden daarop? Is onderzoek naar specifieke zintuigen niet het domein van wetenschappen als psychologie, neurowetenschap, en biologie? Maar als je er over nadenkt, zijn er goede redenen om filosofie te bedrijven over specifieke zintuigen.

De grote vragen in de filosofie kan je opsplitsen tot kleinere vragen. Neem de vraag 'hoe kunnen we betrouwbare kennis verkrijgen?'. Eén zeer invloedrijk antwoord op die vraag luidt dat alle kennis uiteindelijk gebaseerd is op de ervaring. Er zijn verschillende soorten ervaringen en *perceptie* (zintuiglijke waarneming), in tegenstelling tot, zeg, religieuze openbaring lijkt het type ervaring dat kennis kan opleveren. Dus het lijkt er op dat betrouwbare kennis afhangt van de manieren waarop we waarnemen. Als je vragen begint te stellen over de relatie tussen zintuiglijke waarneming en de zaken die waargenomen worden, dan ben je midden in de contemporaine filosofie van cognitieve wetenschappen beland.

Dit proefschrift probeert precies zo'n soort vraag te stellen: naar de relatie tussen zintuiglijke waarnemingen en datgene wat we waarnemen. Ik concentreer me daarbij op een specifiek subtype van waarneming: temperatuurperceptie. Als er in de geschiedenis van de filosofie over waarneming geschreven werd, dan ging het meestal over *visuele* waarneming. Filosofie over visuele waarneming is zeer productief gebleken en heeft veel begrip van filosofische vraagstukken over perceptie in het algemeen opgeleverd. Maar, recentelijk is er binnen de filosofie meer aandacht ontstaan voor niet-visuele zintuigen. In sommige gevallen heeft dat filosofische onderzoek laten zien dat onze manier van denken over waarneming in het algemeen veranderd moet worden om de niet-visuele zintuigen theoretisch te kunnen accommoderen. In dit proefschrift hoop ik iets te doen in de lijn van die hedendaagse filosofie over niet-visuele zintuigen.

Ons zintuig voor temperatuurperceptie is om een paar redenen interessant. Toen ik voor het eerst geïnteresseerd raakte in temperatuurperceptie, dacht ik dat dit zintuig een *eenvoudig* voorbeeld kon zijn (bij wijze van casestudy) voor algemene vraagstukken in de filosofie van de waarneming. Ik dacht dat temperatuurperceptie ons rechtstreeks toegang verleende tot één enkele fysieke grootheid: warmte. Andere zintuigen, zoals *zicht*, geven ons toegang tot allerlei verschillende eigenschappen van de werkelijkheid: denk aan kleurschakeringen, saturatie van kleuren, contrast, afmetingen, vormen, afstanden. Het is een lopend filosofisch en empirisch project om te bepalen *wat het nou precies is* waar ons visuele systeem op is afgesteld. Zo ingewikkeld als de filosofie van visuele perceptie is, zo simpel zou de filosofie van temperatuurperceptie zijn – dacht ik. Temperatuurperceptie is afgesteld om één fysieke grootheid waar te nemen, en deze grootheid is ook nog eens goed gedefinieerd in de natuurwetenschap. Dit 'eenvoudige' zintuig zou als een casestudy kunnen dienen voor meer algemene vraagstukken over zintuiglijke representaties.

Al in een vroeg stadium van mijn doctoraatsonderzoek ontdekte ik dat mijn aannames over de eenvoud van temperatuurperceptie niet klopten. De twee meest prominente stukken uit de recente filosofische literatuur over temperatuurperceptie beweerden allebei dat ons systeem van temperatuurperceptie in feite helemaal geen representaties van temperaturen produceert.[1]

Naarmate ik meer leerde over temperatuurperceptie in mensen en andere dieren, begon ik te begrijpen dat het niet de eenvoud, maar juist de complexiteit van temperatuurperceptie is die het de moeite waard maakt om filosofisch onderzoek te doen. Filosofen hebben historisch gezien zintuigen met name in de bovengenoemde context van de kennisleer bestudeerd: de zoektocht naar betrouwbare kennis over de wereld om ons heen. Maar de evolutie heeft de mens niet uitgerust met eenvoudige thermometers op de huid die betrouwbaar verslag doen van de omgevingstemperatuur. Wat mens en dier wél hebben is een complex temperatuurgevoelig systeem dat erop gericht is om een veilige lichaamstemperatuur te behouden en weefselschade te voorkomen. Qua biologische functie en evolutionaire geschiedenis is ons systeem voor temperatuurperceptie nauw verbonden met het somatosensorische systeem[2]. Temperatuurperceptie heeft een speciale relatie met pijn, en met interoceptie (de zintuigen die onze lichamelijke processen monitoren).

Dit zintuig, dat bij mij in eerste instantie toescheen als een *eenvoudig* geval in de filosofie van de waarneming, is interessant gebleken om andere redenen. De verschillende functies van het systeem voor temperatuurwaarneming hebben nog steeds allemaal te maken met temperatuur, die ene fysieke grootheid waarop de wetenschap een goede grip heeft. Om deze reden is temperatuurperceptie een bijzonder interessant geval voor mensen die geïnteresseerd zijn in filosofische vraagstukken over mentale representaties van fysieke grootheden. Ook al is temperatuurperceptie geen *eenvoudig* geval, het is zeker een *interessant* geval.

In de loop van mijn onderzoek heb ik gemerkt dat een goede theorie van temperatuurperceptie waarschijnlijk indruist tegen sommige geaccepteerde ideeën over perceptie. Specifiek onderzoek ik in dit proefschrift de mogelijkheidsvoorwaarden voor mentale representatie, en ik leg die ideeën langs de meetlat van temperatuurperceptie. Als een theorie van perceptie in het algemeen in strijd is met temperatuurperceptie, dan klopt die theorie waarschijnlijk niet.

Dit proefschrift maakt twee belangrijke claims: de eerste claim is dat mensen een zintuig voor temperatuurperceptie hebben. Dat lijkt misschien voor de hand liggend, maar het is geen universeel geaccepteerde stelling in de filosofie. Er zijn open vragen over hoe je überhaupt kunt vaststellen wat een zintuig is. Wat maakt ruiken onafhankelijk van proeven? Is temperatuurperceptie onafhankelijk van tastzin? Is het los te zien van het somatosensorische

1 Die artikelen waren Akins (1996) en Gray (2013b).

2 Een deel van het sensorisch zenuwstelsel dat gericht is op veranderingen binnen het lichaam en aan het oppervlak van het lichaam..

systeem? Hebben we eigenlijk wel goede criteria om zintuigen van elkaar te onderscheiden? Kan wetenschap ons vertellen welke zintuigen er zijn?

Behalve deze meer algemene vraagstukken over het onderscheid tussen de zintuigen, zijn er ook specifieke filosofische redenen om te denken dat we geen temperatuurzintuig hebben. Eén zo'n reden is het idee dat we niet de temperatuur van onze omgeving waarnemen, maar de snelheid van de uitwisseling van thermische energie tussen ons lichaam en de omgeving.

Een andere positie in het filosofisch debat claimt dat temperatuurperceptie geen *representaties* vormt. Wat dat precies betekent wordt uitgebreid besproken in dit proefschrift, maar voor nu volstaat het om te zeggen dat volgens de meest algemeen geaccepteerde theorie van waarneming de zintuigen representaties produceren. Dus de vraag is: gebruikt ons systeem van temperatuurperceptie representaties?

Dat brengt ons bij de tweede belangrijke claim van dit proefschrift: dat temperatuurperceptie een representationeel vermogen is, en dat het temperaturen representeert. Om die claim te beargumenteren, moet een hoop filosofisch werk verzet worden over het concept van perceptuele representatie. Dit is precies het gebied waarin gevestigde theorieën van waarneming tekortschieten en toe zijn aan een revisie op basis van hun toepasbaarheid op temperatuurperceptie.

Ik beargumenteer dat het concept van mentale representatie vaak wordt verondersteld meer te impliceren dan het daadwerkelijk doet. Bijvoorbeeld: representaties zouden betrouwbaarheid impliceren – iets kan alleen iets anders representeren als ze betrouwbaar samen voorkomen. Of men zegt dat representaties een sterke gelijkenis moeten vertonen met wat ze representeren: in het geval van afbeeldingen, kan je zeggen dat een afbeelding een object representeert wanneer het op het object *lijkt.* Een vergelijkbaar idee bestaat voor zintuiglijke waarneming: waarnemingen kunnen alleen de wereld representeren als er een sterke *structurele gelijkenis* is tussen die twee. Dit is een tamelijk technisch verhaal, maar het wordt op een opmerkelijk 'losse' manier gebruikt in de filosofie. Een goed deel van dit proefschrift is gewijd aan het aanscherpen van deze *vereisten voor representatie.*

Ik beweer dat representaties onderdeel zijn van grotere systemen, die hun eigen functies hebben. De manier waarop zulke systemen informatie over de wereld gebruiken bepaalt zowel of zo'n systeem een representationeel systeem is, als ook de inhoud van die representaties. Deze theorie van representatie legt de nadruk op hoe representaties organismen in staat zijn om bepaalde functies uit te voeren – dit leidt ons weg van het idee dat representaties altijd puur waarachtige objectieve `beelden' van de realiteit zijn.

De representationele theorie van temperatuurperceptie wordt gepresenteerd in de loop van de eerste zes hoofdstukken. In het zevende hoofdstuk wend ik mij tot thermische pijn, en beargumenteer dat er een representationele theorie van pijn bestaat die bijzonder goed past bij pijn veroorzaakt door hete of koude stimuli. In het achtste en laatste hoofdstuk wijk ik enigszins af van de hoofdlijn van het proefschrift en presenteer ik een argument over het concept *levendigheid*, een concept dat soms gebruikt wordt in psychologie om de fenomenologie van perceptie en mentale voorstellingen te beschrijven. Ik denk dat het concept van *levendigheid* onder druk komt te staan wanneer we het toepassen op de waarneming van temperatuur en mentale voorstellingen van temperatuur. Ik heb dit hoofdstuk opgenomen in dit proefschrift omdat het laat zien hoe het denken over temperatuurperceptie een invloed kan hebben op hoe we (zouden moeten) denken over concepten en ideeën binnen de psychologie en de filosofie van de geest, ook als die concepten niet direct betrekking hebben op temperatuurperceptie.

Om samen te vatten, dit proefschrift beargumenteert dat we een zintuig hebben dat temperatuur representeert, en dat het algemene theoretische raamwerk van perceptuele representatie aangepast moet worden om dat zintuig te accommoderen.

Samenvatting

Hoofdstuk 1 gaat over de vraag hoe we zintuigen van elkaar kunnen onderscheiden; het behandelt de vraag van de *individuatie* van zintuigen. Ik presenteer verschillende 'klassieke' theorieën van de individuatie van zintuigen en laat zien dat elk van deze theorieën significante tekortkomingen heeft. Het hoofdstuk gaat verder met contemporaine theorieën van de individuatie van zintuigen, waarbij de problemen die deze theorieën plagen worden toegelicht. De conclusie van het hoofdstuk is dat de beste versies van contemporaine theorieën van de individuatie van zintuigen een soort pragmatisme behelzen: of een bepaald zintuiglijk systeem gezien moet worden als een op zichzelf staand zintuig hangt ten minste ten dele af van het type vragen dat je wilt stellen over dat zintuig.

Hoofdstuk 2 gaat over temperatuurperceptie. Eerst moet er een terminologische kwestie afgehandeld worden, en wordt er een samenvatting gegeven van de empirische wetenschap over temperatuurperceptie. Na deze voorbereidende werkzaamheden, wordt er een argument gemaakt dat de menselijke vermogens om temperatuur waar te nemen beschouwd moeten worden als een zintuig wanneer we een plausibel criterium van de individuatie van zintuigen hanteren. Ik presenteer een verbeterde versie van de neuro-ethologische methode van de individuatie van zintuigen om mijn claim te ondersteunen. Verder behandelt hoofdstuk 2 een filosofisch debat uit de jaren 1960 dat gaat over het contrast tussen het voelen van een lichaamsdeel als koud door het aan te raken, versus het gevoel van een lichaamsdeel als koud zonder het aan te raken.

Hoofdstuk 3 bestaat voornamelijk uit kritiek op het *Heat Exchange Model* van temperatuurperceptie. Dit model houdt in dat onze ervaringen van temperatuurperceptie geen representaties zijn van temperaturen, maar van de snelheid van uitwisseling van thermische energie tussen het lichaam en de omgeving. Ik evalueer verschillende argumenten die gegeven worden voor het *Heat Exchange Model* en concludeer dat deze argumenten niet bewijzen dat dit model beter is dan een simpele theorie van temperatuurperceptie.

Hoofdstuk 4 presenteert een algemene theorie van zintuiglijke representatie. Ik opper een informationele theorie van representatie: zintuiglijke representaties doen wat ze doen omdat ze informatie dragen over de omgeving. De theorie die ik aanhang is geïnspireerd door teleosemantische theorieën over de inhoud van mentale toestanden: de (biologische) functie van een representationeel systeem bepaalt de inhoud van de representaties die het produceert.

Hoofdstuk 5 is een uitbreiding van het idee dat uiteengezet wordt in hoofdstuk 4. Hoofdstuk 5 gaat over structurele representaties: het soort representaties dat informatie draagt over de omgeving omdat het structurele eigenschappen deelt met datgene wat ze representeren.

Structurele representaties kunnen op een andere manier informatie dragen dan non-structurele correlationele representaties.

Hoofdstuk 6 past de theorie van representatie uit de voorgaande twee hoofdstukken toe om zo een representationele theorie van temperatuurperceptie te geven. Deze theorie heeft drie elementen, voor representaties van respectievelijk perceptie van interne lichaams-temperatuur, perceptie van de temperatuur van de huid, en perceptie van de temperatuur van externe objecten. Ik reageer daarnaast op een argument tegen representationele theorieën van temperatuurperceptie. Ik behandel ook een aantal verrassende aspecten van temperatuur-perceptie die leiden tot temperatuurillusies, en ik evalueer in hoeverre die illusies verklaard kunnen worden door de theorie die ik bepleit.

Hoofdstuk 7 probeert thermische pijn te situeren in het landschap van filosofische theorieën van pijn. Ik beargumenteer dat een evaluativistische theorie van pijn bijzonder geschikt is voor duiding van thermische pijn. Meer specifiek hang ik een verklaring van thermische pijn aan die zegt dat pijn een verstoring van homeostase signaleert, en dat zo'n verstoring inherent slecht is. Ik beargumenteer ook dat er een continuïteit is tussen thermische pijn en niet-pijnlijke temperatuurperceptie, waardoor een evaluativistische verklaring van beiden mogelijk is.

Hoofdstuk 8 gaat over het psychologische concept *levendigheid* zoals het toegepast wordt op perceptie en mentaal voorstellingsvermogen. Ik beargumenteer dat dit concept alleen correct toegepast kan worden op temperatuurperceptie als we een bepaalde invulling van dat concept hanteren. Dit laat zien dat onze algemene psychologische termen onderhevig kunnen zijn aan revisie als we temperatuurperceptie in ogenschouw nemen.

Introduction

Temperature perception as a philosophical topic

It might seem odd to write a philosophical thesis on a specific sense. Isn't philosophy supposed to be concerned with 'big' questions, and abstract, a priori answers? Isn't research into specific senses the domain of sciences such as psychology, neuroscience, and biology? But if you think about it, there's good reason to do philosophy about specific senses.

The big questions of philosophy come apart to form smaller questions. Take the question of how we can achieve reliable knowledge: one very influential thought about that question is that all knowledge ultimately derives from experience. There are different kinds of experience, and *perception* rather than, say, revelation, seems like a kind of experience that could provide knowledge. So, it seems that reliable knowledge depends on the ways in which we perceive. Now, if you start asking questions about the relation between perceptual experiences and the things you perceive, then we've landed in the middle of contemporary philosophy of cognitive science.

This thesis aims to ask exactly that kind of question: about the relation between our perceptual states and the things we perceive. To make things more specific, I am focused on a specific subtype of perception: the perception of temperature. In the history of philosophy, writing on perception has often focused on *visual* perception. Philosophical thought about vision has proven fruitful and has brought much insight into philosophical issues about perception in general. However, recently there has been a new philosophical interest in non-visual senses. In some cases, this research has shown us that our way of thinking about perception in general needs revision to theoretically accommodate the non-visual senses. In this thesis I hope to do something in the strain of this current philosophy about non-visual senses.

Our sense of temperature is particularly interesting for a few reasons. When I first became especially interested in temperature perception, I thought that it would provide a *simple* case study for some general issues in philosophy of perception. I thought that temperature perception gives us access to a single physical quantity: heat. Other senses, such as vision, help us perceive a multitude of quantities: think of hues, saturation, contrast, sizes, shapes, distances. It is an ongoing philosophical and empirical project to pin down *what it is* that our visual system is sensitive to. As complicated as the philosophy of vision is, so simple is the philosophy of temperature perception, or so I thought. Temperature perception tracks *one* quantity, and this quantity (temperature) is very well understood in science. This 'simple' sense could be used as a case study for some more general philosophical questions about perceptual representation.

Rather soon in the process of my PhD research, I discovered that my assumption about the simplicity of temperature perception was false. The two most prominent pieces of recent philosophical literature about temperature perception both claimed our system of temperature perception actually *doesn't* represent temperature.[1]

As I began to learn more about temperature perception in humans and other animals, I realized that it was not its simplicity, but rather its complexity that makes temperature perception worth philosophical study. Philosophers in the past have been studying sensory perception mostly in the context mentioned in the beginning of this introduction: the quest for reliable knowledge of the world around us. However, evolution did not equip people with simple skin-mounted thermometers for objectively gauging temperatures in our environment. Rather, it equipped us with a complex system aimed at maintaining a safe body temperature and avoiding tissue damage. In terms of biological function and evolutionary history, the system for temperature perception is intimately tied to our *somatosensory*[2] system. Temperature perception has a special relation to pain, and to interoception (the 'senses' that monitor our bodily processes).

This sense, that at a first glance appeared to me as a *simple case* in the philosophy of perception, has turned out to be interesting for a multitude of reasons. The various functions of our temperature perception system still all relate to *temperature*, that simple single physical quantity of which science has a solid grasp. Because of this, temperature perception is a particularly interesting sense modality for people interested in philosophical questions about mental representations of physical quantities. Although temperature perception is not a *simple* case, it is still an *interesting* case for general questions about perception.

Along the way I have noticed that a good account of temperature perception is likely to be at odds with some generally accepted ideas about perception. Specifically, I explore in this thesis some ideas about what constitutes a mental representation and measure those ideas along the yardstick of temperature perception. If our account of perception in general doesn't accommodate perception of temperature, it's probably false.

I make two main claims in this thesis. The first claim is that humans really have a sense of temperature. This might seem obvious from a commonsense standpoint, but it is not a universally accepted position in philosophy. There are concerns about how to even say what a sense is. What makes smell different from taste? Is temperature perception separate from touch? Is it separable from the somatosensory system? Do we even have good criteria to distinguish the senses from each other? Can science tell us what senses there are?

Besides these more general concerns about distinguishing senses, there are specific philosophical reasons to think we don't have a sense of temperature. One such reason is the idea that we don't really keep track of temperatures in our environment, but rather that we monitor the transfer of thermal energy to-and-from the environment.

Another position on temperature sensation holds that it isn't a *representational* ability. What this means is explored in detail in the thesis, but for now it suffices to say that on the received view of sensory perception, it always involves representation. So, does temperature perception involve representation of temperatures?

---

1 Those articles were Akins (1996) and Gray (2013b).

2 A part of the sensory nervous system which tracks changes inside the body, and at the surface of the body.

This leads to the second main claim of the thesis: that temperature perception is a representational ability, and that it represents temperatures. To argue for that claim, a lot of philosophical work needed to be done on the concept of perceptual representation. This is exactly where general theories of perception are due for revision based on their applicability to temperature perception.

It is my claim that the concept of mental representation has often been taken to imply more than it really does. For example, representation has been said to presuppose reliability: something can't represent something else if it doesn't reliably co-occur. Or, representations are said to have strong resemblance to what they represent: in the case of pictures, a picture is said to represent an object only if it *resembles* the object. An analogous position exists for sensory perception: sense experiences can only represent the world if there is a strong structural similarity between the two. This is a rather technical theory, that has been used in rather loose ways in philosophy. A significant part of this thesis is dedicated to exploring these *requirements for representation*.

I argue that representations are parts of larger systems, with their own functions. The way such systems utilize information about the outside world determines both whether they are representational systems and what the contents of those representations are. This theory of representation puts an emphasis on how representations allow organisms to perform certain functions – steering away from the idea of representations as purely truthful or objective 'pictures' of reality.

The representational theory of temperature perception is developed in the first six chapters. In chapter 7 I turn towards thermal pain and argue that there is a representational theory of pain that works well for pain caused by cold or hot stimuli. Then, in the last chapter, I deviate a bit from the main line of the thesis and make an argument about the concept of *vividness,* a concept that is sometimes used in psychology to describe the phenomenology of perception and mental imagery. I think that the concept of vividness comes under pressure when we try to apply it to the perception and mental imagery of temperature. I have included that chapter in this thesis because it shows how thinking about temperature perception can change our views on concepts and ideas in psychology and the philosophy of mind that are not directly related to temperature perception.

To summarize, the thesis argues that we have a sense modality that represents temperature, and that the overall theoretical framework of perceptual representation should be amended to accommodate that sense.

Summary

Chapter 1 concerns how we can distinguish sense modalities; it deals with the question of sensory individuation. I present several 'classic' theories of sensory individuation and show each of these theories has significant shortcomings. The chapter then turns towards contemporary theories of sensory individuation, highlighting problems they face. The chapter concludes that the best versions of contemporary theories of sensory individuation involve some sort of pragmatism: whether some sensory system should be thought of as a sense modality depends at least in part on the type of questions you want to answer about that sense.

Chapter 2 is all about temperature perception, or *thermoception.* It starts by removing a terminological issue and provides a summary of the empirical science of thermoception. After these preliminaries, it is argued that the human ability to perceive temperature should be thought of as a sense modality, if we adopt a plausible criterion of sensory individuation. I present a modified version of the neuroethological account of sensory individuation to support my claim. Also in chapter 2, I discuss a philosophical debate of the 1960's that centered around the contrast between feeling a certain part of your body to be cold by touching it versus feeling it to be cold *without* touching it.

Chapter 3 consists mostly of a criticism of the Heat Exchange Model of thermoception. This model holds that our thermoceptive experiences do not represent temperatures, but rather represent the rate of thermal energy exchange between the body and the environment. I evaluate several arguments made in favor of the heat exchange model and conclude that these arguments do not prove the heat exchange model is better than a common-sense view of temperature perception.

Chapter 4 presents a general theory of perceptual representation. I propose an informational account of representation: perceptual representations do what they do because they carry information about the environment. The theory I adopt is inspired by teleosemantic accounts of mental content: the (biological) function of representational systems determines the content of the representations they produce.

Chapter 5 expands on the account of mental representation presented in chapter 4. Chapter 5 is concerned with structural representations: the kind of representations that carry information because they share structural features with what they represent. Structural representations can carry information in different ways than non-structural correlational representations.

Chapter 6 applies the theory of representation from the previous two chapters to the case of thermoception to formulate a representational theory of thermoception. This representational

theory has three elements for representations of respectively core body temperature, skin temperature and external object temperature. I reply to an argument against representational accounts of thermoception. I discuss some unexpected features of thermoception that give rise to temperature illusions, and I evaluate to which extent those illusions can be explained by the representational account I propose.

Chapter 7 tries to situate thermal pain in the landscape of philosophical theories about pain. I argue that thermal pain is particularly suited for an evaluativist explanation of pain. Specifically, I endorse an explanation of thermal pain as a perception of a disturbance of homeostasis, which is inherently bad for you. I also argue there is a continuity between thermal pain and non-painful thermoception, and an evaluativist account of both is possible.

Chapter 8 considers the psychological concept of *vividness* as applied to perception and mental imagery. I argue that this concept can only be properly applied to thermoception on a particular understanding of vividness. This shows that our general psychological notions may be subject to revision when temperature perception is taken into consideration.

1 The senses

1.1 Sense modalities

This thesis is about the sense of temperature, or thermoception. In general, perception is a multitude of processes by which animals gain information about their environment to enable them to coordinate their behavior to it. There are many ways in which animals and humans can perceive, and we call these ways 'senses' or 'sense modalities'.

A philosophical theory of temperature perception should include arguments to think whether it is a sense modality or not. In this first chapter of the thesis, I will look at historical and contemporary theories of individuating sense modalities and conclude that they all have shortcomings, some more problematic than others.

Section 1.2 is dedicated to discussion of four classic proposals for the individuation of sense modalities. Section 1.3 is a critical engagement with a contemporary theory of individuation that proposes a non-sparse view of the senses on which a multi-dimensional space of sensory modalities can help us individuate them. Section 1.4 is about the influential neuroethological account of the senses. Section 1.5 concludes that all accounts discussed in this chapter fail to provide a neutral 'metaphysical' way of individuating the senses, and that projects of sensory individuation are best understood as to some degree pragmatic. This view is developed further in chapter 2, where I present my modified neuroethological account and apply it to temperature perception.

1.2 Principles of individuation for sense modalities

In philosophy of mind in the 20th century there has been some discussion on how to answer the question what senses there are (the 'counting question'). In fact, the discussion focuses on what senses *humans* have, a restriction worth noting, as many animals have sensory systems rather different from the ones humans possess. H.P. Grice recognized four ways of individuating the senses that became the onus of discussion. (Grice, 1962) Extensive discussion of these criteria may be a bit tedious to the reader familiar with the literature, as these criteria have been thoroughly discussed and refuted, and most contemporary philosophers tend not to think that any of these four criteria are live options. However, I will still discuss each in turn as combinations of these criteria play a role in contemporary theories on individuating the senses.

Here are the four 'classic' proposals for how to individuate the senses:

1. Senses are individuated by what types of stimuli they respond to.
2. Senses are individuated by what organs or receptors they use.
3. Senses are individuated by the phenomenology of the experience they provide.

4. Senses are individuated by their proper sensibles: the properties this sense responds to.

The first classical proposal under consideration is that sense modalities are individuated by the physical stimulus type they are responsive to. On this view, vision is that sense that is responsive to electromagnetic stimuli within a certain range, audition is the sense that reacts to vibrations in a medium, etc. (Heil, 1983, 2011)

A great deal of the appeal of this proposal lies in that it reduces the question of individuation to an empirical question that can be answered by science; the question what sense modalities there are reduces to the question what stimulus types there are. In contrast to the phenomenological criterion, it doesn't rely on subjective experience to individuate the senses.[1]

A criticism against this view leveled by Keeley (2002) is that physics may give us a space of possible senses, but it doesn't tell us much about what senses there are. It gives us a list of types of energies that sense receptors may be responsive to, but that doesn't tell us if we actually have receptors for those energies.

Keeley's argument doesn't show that the physical stimulus type can't be the difference maker in whether two putative senses under consideration are the same sense or not. For two putative sense modalities we can decide whether they are the same or different by appealing to their stimulus types. Say that putative sense A is sensitive to electromagnetic radiation while putative sense B is sensitive to gravity, then it is seemingly clear they are different senses. But in other cases, the stimulus condition doesn't provide such an easy answer. Take a putative sense modality C that is sensitive to ultraviolet light and a putative sense modality D that is sensitive to radio waves. Both C and D are sensitive to electromagnetic radiation, but the frequency ranges of radiation they are sensitive to do not overlap. Whether or not C and D are the same sense depends on how you individuate the stimulus types: do we consider these non-overlapping ranges of the same energy type to be independent stimulus types? And what about partially overlapping ranges?

It seems to me that the proponent of the stimulus type view has two ways out. The first way is to appeal to physics to tell us which stimulus types are truly distinct. This amounts to a bottom-up approach where physics itself tells us which sense modalities there can be, independent of biological and psychological considerations. Physics doesn't draw any principled boundaries within the electromagnetic spectrum, so ranges of electromagnetic stimulation should not count as different stimulus types. Therefore, C and D are to be counted as the same sense modality. A and B on the other hand are sensitive to forces that are seen by fundamental physics as distinct interactions. It is a standard view in physics that there are four fundamental interactions (forces), namely gravitation, electromagnetism, weak interaction (i.e., weak nuclear force) and strong interaction (strong nuclear force). Physics

[1] For Heil the view is partly motivated by how it deals with cases of perception using sensory substitution devices (particularly tactile-visual sensory substitution). See: Bach-Y-Rita, Collins, Saunders, White, & Scadden (1969).

does draw a principled distinction between gravitation and electromagnetism, so A and B are distinct sense modalities.

Does this bottom-up physical stimulus view mean that fundamentally we could only have four senses? Looking at human sensory systems, we do not seem to possess any that are specifically attuned to weak and strong interactions, even though no senses could exist without those forces. Such a bottom-up approach is not what's defended by the proponents of this view. Rather, the proponents appeal to non-fundamental physical types of stimuli, such as 'vibration in a medium' or 'airborne chemicals'.

My suspicion is that picking out these non-fundamental physical types can only be done by referring to the sense organ or to the properties represented in a sense: physics *per se* doesn't carve out the physical stimulus type 'vibrations in a medium' for hearing. We get to that stimulus type characterization by trying to spell out in physical terms what the *ear* is receptive to. I think the stimulus type account can only be successful when combined with psychological and/or biological considerations. Such blended accounts are discussed in sections 1.3 and 1.4.

The second proposal under discussion is that sense modalities are individuated by the sense organs involved in perception in that modality. On this view, vision is the sense modality that uses the eyes, olfaction the sense modality that involves the nose, etc. A more sophisticated version would be that senses are individuated by the type of receptors involved.[2]

An issue for this view is that it needs to provide a way of individuating *sense organs* in order to individuate sense modalities. A proposal to this effect by D.M. Armstrong says that a sense organ is "a portion of our body which we habitually move at will with the object of perceiving what is going on in our body and environment" (Armstrong, 1968, p. 213). This proposal overlooks the fact that we can't properly move our ears without moving our nose, mouth, vestibular organs, and a good portion of our skin. (Casati, Dokic, & Le Corre, 2014) Moreover, it seems unnecessary that a sense organ be moveable at all. Ways of individuating sense organs that appeal to the stimulus types they are sensitive to, or to the phenomenal character of the experiences the sense organs produce run the risk of collapsing into one of the other classical proposals.

The third proposal is to individuate the senses according to their phenomenal character. The idea is that each sense modality has a distinct type of experience associated with it. On this view, vision is the sense that results in visual experience, audition is the sense that results in the experience of hearing, etc. This view is famously defended by Grice, who thinks the other classical proposals are inconsistent unless they (implicitly) appeal to the phenomenal character of sensory experience. (Grice, 1962).

The view is motivated by a thought experiment that goes something as follows: say we were to encounter Martians that are similar to Earth-humans, but instead of one pair of eyes they have two pairs. Their language does not contain a single verb for 'to see' but rather two

---

[2] The proposal is philosophically unpopular, but it is discussed in much of the literature, so I have included it without reference to any particular authors.

verbs, to *x* and to *y,* where *x* is used to indicate perceiving with one set of eyes, and *y* indicates perceiving with the other set of eyes. The two sets of eyes are both sensitive to the same stimulus types as our earthling eyes, and the properties perceived with both sets are the same as our visual properties. However, the Martians tell us that there is a phenomenological difference between x-ing and y-ing: perceiving something to be F through x-ing feels very different from perceiving something to be F through y-ing. Grice maintains that in such a case only the phenomenological criterion will help us distinguish x-ing from y-ing, while on the other criteria (or combinations thereof) the two Martian senses of x-ing and y-ing would be counted as a single sense, namely *vision.* The correct answer, according to Grice, is that neither x-ing or y-ing counts as seeing.

There is much to be said against this thought experiment. Consider this alternative thought experiment: What if the Martians had only one set of eyes and only x-ed but not y-ed. Arguably, Grice would have to maintain that x-ing is not seeing, since nothing about x-ing is different in this modified thought experiment versus the original. In this situation we encounter beings that do something exactly like seeing with organs similar to ours, but of whom we can't say that they see. The problem is that this scenario of encountering Martians with only one set of eyes is exactly like encountering earth-humans. Since we do not have direct access to other people's experience, it seems to be an unwanted consequence of Grice's thought experiment that we cannot know what senses other earth-humans have. (Coady, 1974)

Or, in another modification to the thought experiment, imagine Martians with one pair of eyes that claim there is a phenomenal difference between using their eyes in the morning and in the afternoon. It is plausible we would be less inclined to accept the idea of two separate senses in such a case: in Grice's thought experiment our intuitions are changed by the fact that there are two sets of sense organs associated, even though that should be irrelevant on the phenomenological account. There is little positive reason to accept two different sense modalities on the basis of reported phenomenal difference alone. (O'Dea, 2011)

Overall, the problem with the thought experiment is what status we attribute to the Martian testimony that x-ing and y-ing are different. Coady says of this:

> Certainly we can usefully refer to sensory experience in some technical discussions, but the Martians seem to be making this notion do a kind of work for which it is very poorly equipped by any of its more acceptable uses. X-experiences are supposed to differ from y-experiences in the kind of way that taste experiences differ from smell experiences, but this "kind of way" is to be understood without reference to differences of organ, medium, or properties. (Coady, 1974)

Setting aside issues with Grice's thought experiments, there are direct reasons to think the phenomenological criterion is not going to give us a good individuation of the senses. Keeley points out that the criterion doesn't allow for sensory perception that doesn't have an associated phenomenology.

> Consider the case of the vomeronasal system. Admittedly, there is still controversy as to whether humans possess this modality, but over the past decade evidence in its favor has begun to mount. Furthermore, if we in fact possess this system, two things about it are striking: first, it plays a significant role in human behavior; and, second, we experience no sensations associated with this modality – there is no "special introspectible character" here, hence no basis to individuate this modality from any other. It would appear to be a modality without sensory experiences. (Keeley, 2002)

If it is the case that we have perception without sensory experience, then the phenomenological criterion would have trouble categorizing those instances of perception as belonging to a sense modality. Besides the example of vomeronasal perception mentioned by Keeley, we could think of more everyday examples of unconscious perception that pose a problem for the phenomenological account.

Another shortcoming of the phenomenological criterion is its inability to explain the surprising phenomenology of the McGurk effect (and other cross-modal illusions). (McGurk & MacDonald, 1976) In the McGurk effect experiment, subjects are presented with a phoneme auditorily, and at the same time they see a video of a person pronouncing a phoneme. Some subjects report *hearing* a phoneme /da/ while in reality the emitted phoneme in the audio component is /ba/, and the phoneme that was presented visually is /ga/. The visual stimulus influences the auditory experience of subjects, while subjects are not aware of this influence. This shows that it may be hard to phenomenally distinguish between sense modalities in particular cases: was it vision or audition that made you hear the /da/? (Casati et al., 2014) The proponent of the phenomenological criterion has an easy answer to this: it felt like hearing, so it was. Although this is not satisfying to those who don't endorse the phenomenological criterion, it is perfectly consistent for those who do.

Finally, there are strong reasons not to base a philosophical or scientific theory on phenomenological self-report. Introspection as a method of reporting on the nature, content or causes of our mental states – though important to philosophy – is highly unreliable. (Bayne & Spener, 2010; Schwitzgebel, 2008)

The fourth and last of the classic theories is that each sense has one or more associated *key features* or *proper sensibles* that can be represented through a particular sense and that sense alone. On this view, audition is the sense that can represent *pitch*, while vision is the sense that can represent *color*. There are also properties that can be represented in multiple senses, such as number or magnitude – these are the *common sensibles*. This view was espoused by Aristotle and finds more recent support from Roxbee Cox. (1970) Key features of sense modalities are properties that can be *directly* perceived (without perceiving another property). For example, shape is not directly perceived in vision, because we cannot see shape without seeing color. Color however can be perceived without seeing a shape. Thus, according to Roxbee Cox, color must be the key feature of vision.

A problem with this account is, it is not always clear that these key features can be perceived by *only* one sense modality. We can imagine cases where a device translates color information into auditory information. In such a case, do we hear colors, or do we see them?

Whatever the answer might be to this question, it seems to be a contingent fact rather than a necessary truth that colors are detectable only by vision. (Heil, 1983)

The challenge for the proponent of the key feature view is to identify the key feature without reference to the sense modality it aims to define. If that is impossible, then definitions of sense modalities will be circular. But for key features that are identified without reference to the sense modality, the question remains what is *essential* about these features to the sense modality. (Casati et al., 2014)

1.3 A space of sensory modalities

Contemporary authors have continued the discussion of individuating the senses in more promising directions. One interesting proposal by Fiona Macpherson is to be very liberal in what information can inform the project of distinguishing the senses. On this account, a principle of individuation is not pre-decided, but rather information about different senses is said to carve out a *space* of sense modalities. According to Macpherson, we can take the four classic criteria and from them construct a multidimensional space in which we can plot sense modalities. A cursory version of this view can be found in Macpherson's (2011b and 2011a), which garnered criticism in Gray (2013a). In response to this criticism a more fully fledged view is presented in Macpherson (2014).

Macpherson's account is motivated by criticism of the four classical criteria of individuation. According to Macpherson these classical criteria result in 'sparse views' which recognize only a limited number of senses. Macpherson's criticism is that each of these criteria fail when confronted with non-standard senses such as: senses that fall outside the five Aristotelian senses, modified Aristotelian senses, malfunctioning Aristotelian senses, and Aristotelian senses outside of their usual environments. For example, Macpherson argues that none of the classical criteria do well at categorizing bee vision, or vision using a tactile-visual sensory substitution device.[3]

In response to the failure of the classical criteria to deal with non-standard senses Macpherson proposes a non-sparse view of the senses, on which the senses may not be always rigidly delineated and on which there are many more different senses than the mere five recognized by Aristotle. A multi-dimensional space helps us make sense of the senses by showing us the relations between these many senses. On the early version of Macpherson's view, it reads like the four criteria are to constitute four dimensions of the space:

> I hold that the four criteria are relatively independent dimensions along which different possible kinds of senses could take different values. We can think of these

[3] A tactile-visual sensory substitution (TVSS) device is a device that translates visual information (from a camera) into tactile stimuli that are administered through a wearable mechanism. (Bach-Y-Rita et al., 1969) These devices, after training, can enable persons who are visually impaired to 'visually' navigate environments including obstacles.

> four criteria as defining a multidimensional space within which we can locate each of the Aristotelian senses, the four examples of unusual senses discussed earlier, and any other sense. (Macpherson, 2011a, p. 140)

Later Macpherson clarifies that the modeled space does not consist of four dimensions corresponding to the four criteria, but rather consists of many more dimensions which are informed by but not identical to the criteria. (Macpherson, 2014)

The idea is that the sensory space will provide a way of organizing information about the senses that allows us to compare the putative senses. The basic principle is that senses that are more alike will be closer together (or overlap more) in the space of sensory modalities. For example, human vision and bee vision would be closer together and therefore more similar than human vision and bat echolocation. In MacPherson's view, plotting senses in the multidimensional space and applying measures of similarity would provide a fine-grained taxonomy of the senses that is ultimately more informative than any of the four classic criteria. At the very least it should allow us to say some sense A is more similar to sense B than it is to sense C. Adding to that, if putative senses form close clusters in the space of sense modalities, we may want to say they are the same sense. The resulting view is a *rich* view of sensory modalities, on which there is not a limited number of modalities but many, and modalities need not be totally distinct from one another.

The idea of *combining* the criteria in subtle ways is interesting, and the versatility of Macpherson's account is an attractive feature. However, working out a space of sensory modalities as Macpherson proposes is rife with implementation problems, that go directly to the heart of the matter. I will first review Gray's arguments against Macphersons early cursory presentation of the view, and then add my own arguments against the more full-fledged version of it.

Non-ordering problem

If we take Macpherson's early view to be that the four classical theories *are* the axes of the space, we run into trouble quickly. The criteria simply do not form *axes* in any straightforward way, because they don't form non-arbitrary orderings. The proximal stimulus dimension does not form an ordering because there is no non-arbitrarily order in e.g. {chemical stimuli, vibration in a medium, electromagnetic radiation}. Similarly, it is unclear how different sense organs should be ordered if we want a non-arbitrary similarity relation to hold: is a nose more like an ear than an eye is? The same problem holds for the key-feature axis and arguably the phenomenological axis. The different possible values on each of these 'dimensions' simply do not form a natural ordering. (Gray, 2013a)

No value on X problem

Another problem lies in that some senses one wants to plot do not naturally take a value on certain dimensions. Say we have a dimension for the frequency of soundwaves a sense is receptive to. Human hearing and bat hearing occupy different but overlapping ranges of this dimension. Human vision, however, is not sensitive to sound waves at all. If vision does not take a value on this dimension, then the space in which we've plotted it is not the same space in which we've plotted hearing.

Similarity between disparate senses

Because the dimensions of the space are joined at some origin, it can turn out that things we think of as very different senses are closer in the space than things we think of as the same sense. In the figure below we can see a hypothesized space in which hearing is more similar to low-wavelength vision than low-wavelength vision is to high-wavelength vision.

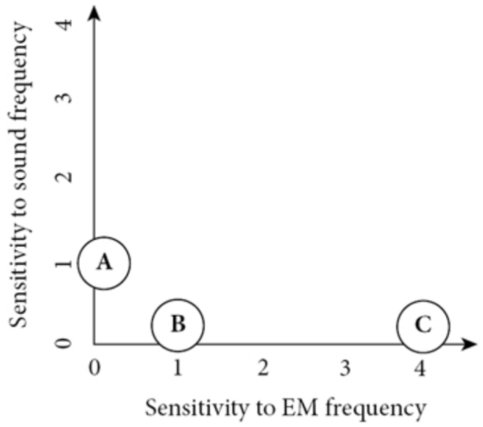


Figure 1: Low-wavelength vision (B) may be closer in the space to low-wavelength hearing (A) than it is to high-wavelength vision (C). Image from Macpherson (2014).

In response to Gray's objections, Macpherson has proposed an alternative form of the dimensional view, with other dimensions that lend themselves to constructing a space. (Macpherson, 2014) On the revised view, the four classical criteria are not taken to constitute four dimensions of the space of sensory modalities, but rather *inform* which dimensions there could be. To take the *stimulus type* criterium, Macpherson's suggestion is that we take dimensions corresponding to energy types, and separate dimensions for the sensitivity to that energy type. For example, we take a dimension for wavelength of electromagnetic radiation, and a dimension for sensitivity to electromagnetic radiation. These dimensions have a natural ordering (from low to high wavelength and from low to high sensitivity) and are therefore not subject to the non-ordering problem.

This approach also helps with the *no value on x* problem: senses that do not have a sensitivity to the electromagnetic spectrum can be plotted as having a 0 sensitivity to electromagnetic wavelength, as seen in figure 2. In this way, all senses can be plotted in the same space.

In response to the problem of similarity between disparate senses, Macpherson argues that what's important is the similarity of certain senses *with respect to specific dimensions.* Macpherson says that "*If* the senses can be plotted in the space, that is, *if* they take a value on both dimensions, *then* their place in this space reflects their similarity with respect to these two values." (Macpherson, 2014, original emphasis) On the picture above in figure 1, if *A* really takes a value of 0 on the x-axis, then *B* is more similar to *A* than to *C* with respect to the x-axis. According to Macpherson however, figure 1 shows wrong dimensions. Sensitivity to wavelengths is a dimension on which hearing cannot be plotted. Rather we should choose to include separate wavelength and sensitivity dimensions, as shown in figure 2.

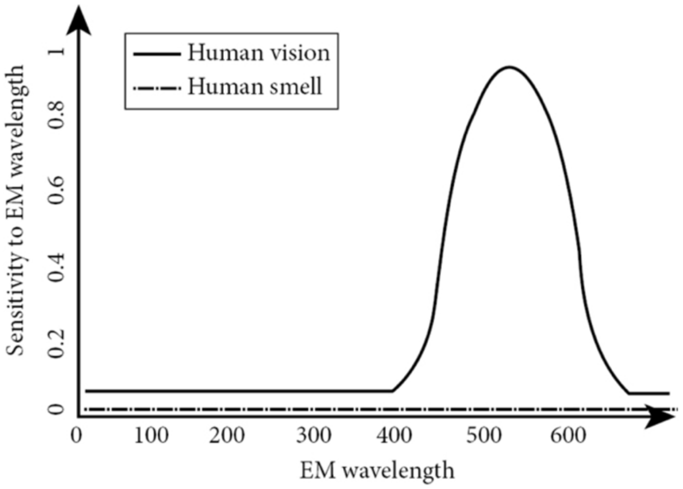


Figure 2: A two-dimensional space that plots human vision and human smell for their sensitivity (normalized to a peak value of 1) against wavelength of electromagnetic radiation). Image from Macpherson (2014).

Although Macpherson's response to Gray is adequate, I think problems of similarity of disparate senses persist, even when the space is constructed as Macpherson suggests. Take a space with the following four dimensions:
x: EM wavelength
y: EM sensitivity
z: Sound wavelength
w: Sound sensitivity

Now imagine the following three senses plotted in the space:

1. Low-wavelength vision: only sensitive to short EM wavelengths (near the origin of the space), not sensitive to sound.
2. High-wavelength vision: only sensitive to very high EM wavelengths (far from the origin of the space), not sensitive to sound.
3. Low-wavelength audition: only sensitive to low wavelength sound (near the origin of the space), not sensitive to electromagnetic radiation.

Sense 1 takes the form of a blob on the x and y axis near the origin but is a line at 0 sensitivity through the z and w axes. Sense 2 takes the form of a blob on the x and y axis, but far away from the origin, and is a line at 0 sensitivity through the z and w axes. Sense 3 takes the form of a blob on the z and w axis near the origin and is a line at 0 sensitivity through the x and y axes.

Although it is a bit hard to visualize a four-dimensional space, we may think of these three senses as regions in the space defined by {x, y, z, w}. Low-wavelength vision and low-wavelength audition inhabit an area close to the origin of the four dimensions. High-wavelength vision inhabits an area further from that origin. Senses 1 and 3, which are close to the origin, could be closer together than 1 and 2 if we calculate the distances between these senses in four dimensions. If we understand the similarity of senses in terms of distance in this space, then it turns out low-wavelength vision and low-wavelength audition are more similar than low-wavelength vision is to high-wavelength vision.

Macpherson seems to suggest this is not a problem because what matters is the similarity between two senses *with respect to a dimension*. If I understand this claim correctly, Macpherson is saying that the fact that sense 1 and 3 are closer together in our four-dimensional space than sense 1 and 2 doesn't really mean anything. If that is the case, we are left wondering why you would construct a space in the first place. Plotting data as points or lines or regions along axes that correspond to variables is usually done with a view to see similarity and dissimilarity between the data points not only regarding a single dimension, but the relation between those dimensions. If the distance between the regions in a space is not to be the individuating principle for the plotted senses, then what is the use of plotting these senses in that space?

An option that is open to the dimensionalist in response to this objection is to simply allow that sense 1 and 3 in the example above are more similar than sense 1 and 2. The dimensionalist account from the outset was thought to provide a non-sparse view and would likely be quite revisionary.

Puzzle of selecting dimensions

Even if we accept that dimensionalism about the senses is going to result in a very revisionary categorization of the senses, a worry remains. When we set out to plot the putative senses in the space, we must decide which dimensions to include. I claim that the outcome of the dimensionalist method critically depends on this selection of dimensions, and that circularity looms large: if the dimensions are selected to confirm our pre-theoretic ideas about sensory individuation, then we haven't learned anything.

On the dimensionalist account, much depends on what dimensions you construct the space out of. For example, including a dimension for wavelength of electromagnetic energy would help distinguish between the vision of *Melanophila acuminata* and human vision, as this species of beetle is sensitive to radiation of infrared wavelengths, while human vision is not. Therefore, if we include the wavelength dimension in our space, the areas covered by human vision and *M. acuminata* vision will not fully overlap. Introducing this dimension gives us grounds to say beetle vision and human vision are not the same sense. If we were to exclude it, this piece of information would not be represented in the theory.

A worry that arises from this fact is that the pre-theoretical choice of which dimensions to include will influence the outcome of the analysis on the space. These pre-theoretical choices would probably be made on an intuitive basis. If this is the case, creating a space of sensory modalities will mostly be an exercise in mapping our intuitions, rather than an enterprise revealing new insights on the individuation of the senses. The worry is that the method for individuating the senses leads to circularity: whatever we put into our method is what we get out of it.

The dimensionalist method has a proposed solution to this issue. Using the statistical tool *principal component analysis* (PCA), you can reduce the number of dimensions in a space by excluding dimensions that do not make much difference to the shape senses take in the space. What PCA does is in this context is simplify spaces: it reduces the number of dimensions in a space and retains only those that do most of the difference-making in the data. In this way you eliminate dimensions that are not of importance for the individuation of the senses. Because PCA has no limit to the number of dimensions involved, other than the computational capacity of the computer running it, you could initially select a large number of dimensions to construct the space and then use PCA to reduce to important dimensions only. In this way we can include not only the dimensions we find intuitively attractive, but many others as well. The dimensions that do not make much of a difference are then rooted out through PCA. So, the worry of circularity is alleviated by the high number of dimensions that can be included in the space.

A stronger problem of selection of dimensions would arise if we could show that there are dimensions that we find intuitively unimportant to the individuation of the senses, but that do result in a big spread of the data in the space. Then PCA would not eliminate these dimensions. Think of the following possible dimensions that might have exactly such an effect: the concentration of staphylococcus aureus on the sense organ, or the relative amount of carbon in the sense organ.

The first of these dimensions would effectively distance human retronasal perception from human perception of heat. The organs for retronasal perception are located in the nasal cavity, which usually hosts a population of *Staphylococcus aureus*. Our heat receptors are mostly unmyelinated C-fibers located in the epidermis, under the surface. We are much less likely to find *S. aureus* there. These bacteria may also occur in the middle ear as they are transmitted there through the eustachian tube. Thus, we can assume for the sake of argument that the concentration of *S. aureus* on the vestibular sense organs, which are located in the middle ear, is more like the concentration in the nose than like the concentration in the skin. So, the dimension gives us an ordering of the senses, dependent on a physical criterium which brings out differences and similarities between the senses.

Intuitively it doesn't seem like we would want to use this dimension to individuate the senses. We could imagine an individual who had topical antibiotics applied to the inside of the eustachian tube, effectively preventing *S. aureus* from entering the middle ear, while still having the bacteria in their nose. Although *S. aureus* is not involved in the function of the vestibular senses, the vestibular senses of this person would still occupy a different region of our space than the vestibular senses of another person, without the antibiotics, would. Thus, it seems we have hit on a dimension which is intuitively spurious since it has nothing to do with the functioning of the sense modality but would still play a role in individuation were it to be incorporated in the space.

Now consider the dimension that orders organs by the relative amount of carbon atoms in them. This dimension could have a similar effect of spurious individuation. Let's say an eye consists of 18.5% carbon (the average for the human body). Now take the eye of someone who had cataract surgery, in which the eye's crystalline lens was replaced with a silicone intraocular lens. This changes the concentration of carbon. In other respects, the eye is very much like a normal eye. Therefore, the carbon-concentration dimension would pull apart two senses that we intuitively think of as the same sense. The normal eye and the post-surgery eye would be the sense organs for different senses since they occupy different regions in the multidimensional space.

Or, to take a different example that leads to spuriousness on this dimension, think of cochlear implants used as neuroprosthetic devices for persons with sensorineural hearing loss. This device takes over many of the functions of a normal ear, but with very different 'hardware'. The cochlear implant would contain a different amount of carbon than the parts of the ear that would normally be involved in audition. Although hearing with a cochlear implant is different from typical hearing,[4] the relevant difference making factor from a functional standpoint hardly seems to be the concentration of carbon in the implant.

Now each of these spurious dimensions on their own are not very problematic for the dimensionalist. Mcpherson has argued for a non-sparse account of the senses, so 'pulling apart' into two senses a sense that we intuitively regard as unified is not problematic for the dimensionalist account. Crystalline-lens-eye vision and silicon-eye vision could be two senses that are very similar but differ in what they are made of. But what is the consequence if we add not one or two or three of these kind of dimensions, but many more? If too many of these intuitively spurious dimensions are added to the space, the senses will be individuated more by the spurious dimensions than by the intuitively plausible dimensions. If the spurious dimensions are sufficiently influential on the shape the data takes in the space, reducing the dimensionality by PCA might even get rid of intuitively plausible dimensions in favor of these spurious ones. This is an effect of the fact that there is no selection procedure for dimensions.

A dimensionalist must either accept this unsatisfying consequence or must say that there is a selection procedure for dimensions. This would, however, raise the problem of circularity again. For what could inform this selection of dimensions other than the kind of considerations that are given for the canonical answers to the question of individuation?

A possible answer for the dimensionalist would be that the four classical criteria pick out aspects of sensory perception that matter to us when we talk about the senses. These aspects of sensory perception are what we are interested in when we research them or when we use concepts of sensory modalities in everyday life, and this is why we are justified in selecting dimensions on the basis of these criteria. But to give this answer would be a considerable defeat to the dimensionalist since it is tantamount to admitting that dimensionalism is merely a way of structuring intuitions based on the classical criteria, and not a self-contained or theory-neutral way of individuating the sense.

The problem of scaling or weighting

The results of PCA are dependent on the scaling of the axes. Since PCA strives to retain maximal variability in the space, 'longer' dimensions are favored. If we had a dimension that represented length in our space, whether we scaled it in millimeters or miles would make a huge difference on the spread of the data over the dimension. In millimeters, the spread would be much larger, and therefore the axis would be seen as more important for the PCA

---

[4] Learning to 'hear' with a cochlear implant requires extensive training. Language development in deaf children with cochlear implants is different from hearing children.

than if it had been miles. The relative influence of the axis in respect to the other axes changes when you change the scaling without changing the scaling of the other axes. It is a known limitation of PCA that, if the units and scaling of the various axes are not the same, the outcome of the analysis is to some degree arbitrary. One way this problem could be alleviated is by giving the different axes different 'weight', that is, assigning different levels of importance to the axes to compensate for the scaling issue. This is not hard: if you wished to reduce the importance of an axis by a factor 8 one could simply divide every value on that axis by 8. This would have the same effect as rescaling the axis. A more sophisticated way of compensating for the scaling problem is to scale all dimensions to *unit variance* (UV). In this procedure the highest value on each variable is scaled to be one, and the lowest zero. All values of all variables then fall between zero and one. Through this procedure each dimension gets the same length. If one wanted to equalize the dimensions even more one could assign each dimension a weight of one divided by the standard deviation on that dimension. In this way, even if we have dimensions of equal length, the spread of the data over that dimension still influences the outcome of PCA.

Although scaling to UV and weighting based on standard deviation gives us a better way to structure our dimensions for PCA, I see no reason to accept they give us a better way to individuate the senses.

Firstly, it seems somewhat arbitrary to say that all these very different dimensions have the same distance between the lowest value and the highest value. Why can't we say that there really is more differentiation going on in one dimension than in another? Admittedly, comparing distances on unrelated scales is impossible, but that doesn't mean it's entirely senseless to want to have one dimension doing more work in the individuation than another based on the range of the dimension. Our intuitions on this might conflict with the outcomes of UV-scaled PCA.

Secondly, some dimensions may use the same units but have different ranges.
Consider the following chart: it plots the height and width in centimeters of four objects.

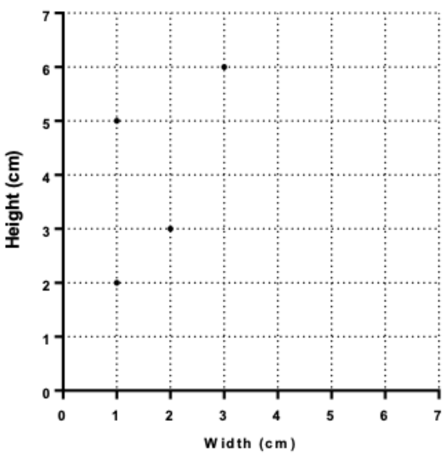


Figure 3: A scatterplot of the height and width of four objects.

From this image you can determine that all objects are vertically oriented oblongs: the height of every shape plotted is larger than the width. If we were to scale this data set to UV that information would be lost. Consider for example the object [H6:W3]. This object is both the tallest and the widest, so when scaled to UV it would become [H1:W1]. In this case both axes are the same units (cm), so it doesn't make sense to scale them to UV. The plotted points can be compared better if they are not scaled to UV. Possibly, some of the dimensions in our multi-dimensional will share the same units. We might want to retain the comparability this affords us. In that case scaling to UV would not be a good idea.

In the end, scaling to UV gives a different outcome of PCA than not scaling to UV. Because of this, scientists sometimes do both UV-scaled PCA and PCA over data that hasn't been rescaled. However, if we were to take this approach to individuating the senses, we would bring back in the original problem of scaling. Because of these issues I think it is contentious that scaling to UV and weighting by standard deviation is objectively the best way to weigh and scale the data in this case. Then should we determine the scaling and weighting by means of pre-theoretical considerations? That could be a solution, but then a worry of circularity similar to the one above arises. For if we choose to assign a larger weight to the axes that we pre-theoretically find to be more important, aren't we just mapping our intuitions about the senses, rather than individuating them by independent criteria?

To summarize the problems with dimensionalism, we can say that it is likely to yield outcomes in strong conflict with our pre-established notions of the senses. We can limit our selection of dimensions to those that will provide an ordering, but which dimensions are selected can have consequences for how the senses are individuated. Using PCA we could narrow down the space to only dimensions that explain the spread of the data, but the outcome of this process may again be in stark contrast to what we think are important aspects of the senses. Implementation puzzles regarding the scaling and weighting of axes in the space are not mere technical issues: they can have consequences for the outcome of the dimensional method. Therefore, the dimensional view does not provide a 'neutral' way of individuating the senses: rather, you get out of it what you put into it.

1.4 Sense modalities as bundles of capacities

Recently a view has been developed by O'Callaghan that sees sense modalities as bundles of information-gathering capacities. Senses, according to O'Callaghan, are distinct ways in which an organism can gather information about the environment. Each sense is a collection of sensory capacities. Sensory capacities are stable psychological traits that are dispositional in nature: the *actual* exercise of a capacity is not essential. To possess a perceptual capacity, according to O'Callaghan, is to be differentially sensitive to things or features of the environment – and for that sensitivity to play a role in further psychological processes such as memory, mental imagery, and judgement. This theory combines two important views on sensory perception: first, that sensory perception is information gathering. Second, that senses are *ways* of perceiving. (O'Callaghan, 2019)

So, according to O'Callaghan, a sense modality is a bundle of differential sensitivities to environmental features that share the *way* in which they gather information. This definition, in and of itself, does not individuate sense modalities: it must be supplemented with a theory of *ways of perceiving.* The idea of sense modalities as bundles of capacities is in fact compatible with the classical criteria: vision, for example, could be defined as the set of capacities that the eye grants an organism (individuation by sense organ). Under this definition of vision, the shared way of perceiving would be that all visual capacities gather information by way of the eye. Neither the insight that senses are information gathering systems, nor the insight that senses are ways of perceiving directly provides a way of individuating the senses.

O'Callaghan argues that sensory capacities are bundled into sense modalities by their *information extraction function*. An information extraction function is the function of a perceptual activity (or capacity) to extract a certain type of information from a certain medium. For example, the perceptual activity of *seeing colors* is defined by its function to extract color information from light (the medium).

This is a good start for a theory of individuation: it distinguishes, for example, vision from hearing by the medium (light vs. sound waves). But it also leaves many cases open. For example, color vision in trichromatic humans (the most prevalent type of color vision) has more differential sensitivity to colors than color vision in humans with protanopia (severe red-green color blindness). When a trichromat and person with protanopia are engaged in the perceptual activity of seeing colors, are they exercising the same sensory capacities? The answer to that depends on how finely the information extraction functions are individuated. If we say that the function is *extracting color information from light,* then they are engaged in the same activity. But in a more fine-grained way of individuating functions, they are not engaging in the same activity, since the trichromat can distinguish green from red, while the person with protanopia cannot. They are both extracting information from the same medium, but they do not have the same capacities as to what information they can extract.

Not only is there a question of grain in the 'information' half of the information extraction function, but there is also a similar issue with the individuation of a medium. Between animal species, there is great variation in the part of the electromagnetic spectrum that animals are sensitive to. Take the infrared sensitivity of *Melanophila acuminata* – when this beetle is seeing infrared color, is it engaged in the same activity as I am when I am seeing some color that falls in the humanly visible spectrum?

To O'Callaghan, indeterminacy is not so much a problem with the theory as it is a feature of it. He says: "One virtue of this schema is that it allows for some flexibility in individuating and counting senses. Differing explanatory projects have differing purposes." (O'Callaghan, 2019, p. 162) The grain at which we may want to individuate senses in a scientific context is perhaps different from the grain at which we want to individuate senses in everyday conversation. O'Callaghan's theory provides an individuation scheme, but not a single answer to the question which senses there are.

1.5 The neuroethological account

Perhaps the most widely accepted contemporary view on individuating the senses is Brian Keeley's idea that:

> to possess a genuine sensory modality is to possess an appropriately wired-up sense organ that is historically dedicated to facilitating behavior with respect to an identifiable physical class of energy. (Keeley, 2002)

Keeley's view is really a combination of the sense-organ view and the stimulus type view. It does not rely on questionable phenomenological introspection, and it does not face the problems of the key feature account. The view is aimed to align itself with the way neuroethology classifies sensory modalities – it is meant to provide theoretical foundation for a scientific practice of individuating sense modalities.

Keeley's account comes down to a set of four criteria that are said to be jointly sufficient and individually necessary for something to count as a sense modality.

> (1) Physics: the external physical conditions upon which the senses depend. That is, we might distinguish the senses by reference to the physical qualities of their respective stimuli: vision is the detection of differences in electromagnetic stimuli; olfaction is the detection of differences in concentration of chemical stimuli. (Keeley, 2002, p. 12)

As noted above, Keeley sees the physical criterium as providing a space of possible senses – what types of energy there are that senses can be responsive to.

> (2) Neurobiology: the character of the putative sense organs and their modes of connection with the brain. For example, vision is what we do with our eyes; audition is what we do with our cochlea and associated brain areas. (p. 13).

The sense organs and their connection to the central nervous system lie at the heart of the view. But (1) and (2) together don't exclude some problematic cases, hence the inclusion of two more criteria.

> (3) Behaviour: the ability to discriminate behaviourally between stimuli that differ only in terms of a particular physical energy type. (p. 14).

This third criterium is included by Keeley in order to rule out vestigial sense organs (e.g., the human vomeronasal organ) that have the appropriate physical structure but do not play a role in the behavior of the organism.

> (4) Dedication: the evolutionary or developmental importance of the putative sense to an organism. For example, we ought not attribute an electrical modality to an individual unless electrical properties of the world are part of the normal environment of that individual and to which the organism is attuned. (p. 17).

This fourth criterium is meant to rule out cases such as the sensitivity of the ear to electric shock, the eye to pressure, etc. In effect the criterium differentiates detection from reception: an eye detects pressure, but it is only receptive to light.

A criticism one can have of Keeley's account is that what *it* classifies as a sense modality isn't really what we ordinarily call a sense – the account is an account of sense modalities *in* *science*, not an explication of the commonsense conception of the senses. (Nudds, 2004) According to Nudds the project of individuating sense modalities for science is a different one from explaining why we have the commonsense conception of the senses that we do. Nudds argues that a scientific account of the senses is not an account of the senses as we commonly understand them – rather than offer an explanation, we've simply changed the subject. If the scientific account is meant to replace the commonsense conception, then we need to offer reasons that "common-sense embodies the kind of proto-scientific understanding of the senses which is *liable* to revision or replacement." (Nudds, 2004, p. 35)

Nudds' argument is that common-sense sense terms play a role in understanding ourselves and others. To know that Alice *saw* the vase is more informative than to know that she *perceived* the vase. And if I know that Alice *felt* the vase rather than *saw* it, that tells me something about the kinds of properties of the vase Alice might have perceived: she can know its shape, but not its color. Nudds thinks that this is what makes the concept of a sense significant to us, and that is why a classification of the senses must respect these *ways of perceiving* that we take ourselves to be employing.

I am not convinced that a scientific account of the senses couldn't provide the kind of information Nudds thinks is required in everyday interaction. For example, echolocation is not one of the commonsense senses, but some people are adept at navigating environments on the way self-produced sounds reflect off surfaces. (Thaler & Goodale, 2016) To know that Alice echolocated the couch is more informative than merely knowing she perceived it. And if I know Alice perceived the couch through echolocation, that tells me something about the kinds of properties of the couch Alice might have perceived: she can know its size and density, but not its color. While echolocation may not be in our repertoire of common-sense ways perceiving, it seems like it could be. It seems to me that a scientific theory of echolocation that provides us with an account of how it works, what properties it represents, and what stimuli it is and isn't receptive to could in fact aid our understanding of ourselves as perceivers.

Still, even if we grant Nudds the point that a scientific account of the senses has a different explanandum than the commonsense senses, for the project of this thesis a scientific way of individuating senses is more apt, since I am trying to give a philosophical theory of thermoception, not necessarily a common-sense theory of thermoception.

Two counterexamples to Keeley's theory have been posed by Gray (2005). The first counterexample revolves around a sensory system of pit vipers or *crotalinae*. These snakes have a pair of heat-sensing pits located on the sides of the head. The pits are lined with sensory endings that are receptive to electromagnetic radiation in the infra-red range. The structure of the pit lets the organs function like a lens-less camera, giving them a putative sensory modality known as *heat imaging*. The vipers use these organs in locating mammals – either as prey or as threat. (Bullock & Cowles, 1952) Besides heat-sensing pits, pit vipers also have eyes sensitive to a different part of the electromagnetic spectrum. Since physics draws no principled boundaries within the electromagnetic spectrum, it is the case that viper pits and eyes respond to the same physical energy type. Since Keeley contended that (1) was an individually necessary condition on sensory individuation, it cannot be the case that *heat imaging* and *vision* are different sensory modalities. Now this is at odds with both scientific practice (neuroethologists tend to see heat imaging as a separate sense) and with common sense. (Gray, 2005)

The second counterexample is the inverse of the first. In pit vipers there are two organs dedicated to a single type of physical energy. The second counterexample deals with the leaf-shaped nose of common vampire bats (*Desmodus rotundus)*, that is receptive to two kinds of energy. Vampire bat noses have thermoreceptors that are used to locate capillaries under the skin of their prey. The receptors are responsive both to conductive heat and radiated heat. Conductive heat is a type of kinetic energy, while radiated heat is a type of electromagnetic energy. So, physics tells us, there are two possible sense modalities (for radiated heat and for conductive heat). Going by Keeley's (1), vampire bats must have two separate sensory modalities located in the same organ. Given that there is only one organ as outlined by (2) and no discrimination between the two energy types in behavior (3) or dedication (4), to say that there are two modalities is unsatisfactory. It runs counter to the scientific practice in neuroethology and to common sense. (Gray, 2005)

The human sense of temperature in fact causes similar trouble for Keeley's account. It uses a variety of receptors distributed through the body to form representations of temperature. And just like the thermoreceptors in Gray's vampiric bats, our thermoreceptors are sensitive to both radiant heat and conducted heat. This is simple physics: all matter at temperatures above absolute zero emits thermal radiation, and thermal radiation will heat up any matter it strikes. Thermoreceptors and tissue surrounding thermoreceptors can be heated with radiation – so all thermoreceptors are in principle sensitive to radiant heat as well as conducted heat.

The two counterexamples point out that to get a scientifically plausible theory of individuating senses from Keeley's four criteria, we need to re-evaluate how the four criteria combine to form a principle of individuation. According to Gray, neuroethology should not apply criteria (1)-(4) as individually necessary and jointly sufficient, but rather should look on a case-by case basis. We may have to judge senses receptive to different ranges of the same energy type as different senses, based on the sense organs involved, the evolutionary history of said organs, and behavioral abilities associated with them. (Gray, 2005)

To put it differently, Gray thinks that when criteria (2)-(4) give a clear picture of a single sense modality, then we may relax criterium (1) in one of two ways: either a single energy type can be split into ranges so as to allow for two senses of one energy type, or we may allow for a sense that is responsive to multiple energy types. This reflects the scientific practice of individuating the senses.

Matthew Fulkerson raises the special case of the sense of touch as problem for Keeley's account. (Fulkerson, 2014) The sense of touch uses a variety of distinct transducers with different physiological structures, different evolutionary histories and different neurological channels. (Lumpkin & Caterina, 2007) As such, touch fails the requirement (2) for having a single 'appropriately wired-up sense organ'. If we strictly applied Keeley's view to the sense of touch, we would have to conclude there is a whole series of senses of touch – one for each type of transducer.

Touch is not the only sense that uses a variety of transducers. Taste receptors come in different kinds that are receptive to different chemical stimuli and there is a variety of photosensitive cells in the retina. These receptors for taste and vision, just like the ones for touch, didn't all evolve simultaneously or as a result of the same environmental pressures. In the case of taste and vision, the receptors are clustered on the tongue and in the retinas respectively.

This is not the case for touch: dermal receptors are spread out over the body. To say that the skin is the sense organ for touch would also be wrong, since touch also uses receptors in muscles, joints, and tendons. (Lederman & Klatzky, 2009) So even if we do not understand Keeley's (2) to imply a single receptor type, it is hard to say exactly what the sense organ for touch is. This difficulty is exacerbated by the fact that many of the transducers have multiple functions (e.g., nociception *and* touch).[5] The skin as a whole also has more functions than just as a sense organ.

Fulkerson contends that "what is missing from Keeley's view … is a role for the *way* distinct transducer populations work together, coordinating their operations into a unified, coherent sensory representation". (Fulkerson, 2014) Fulkerson sees this as a reason to abandon the idea of neatly individuating senses, rather opting for a 'multisensory perspective' of perception as the work of "deeply connected, highly interacting, and mutually influencing sets of sensory systems". (Fulkerson, 2014) This leads to pragmatism about the individuation of sense modalities. It depends on the questions one is trying to answer whether it is useful to see touch as a single sense, or as multisensory. This is not unique to touch; depending on the levels of explanation, other senses (e.g., vision) can be seen as multisensory. According to Fulkerson, there is nothing uniquely multimodal about touch. On the criteria for multimodality where touch turns out to be multimodal, other senses are multimodal as well.

In answer to Fulkerson's objection, we may try to save the Keeley account by making a move similar to Gray's. Where Gray grapples with the status of (1), Fulkerson's objection attacks (2). Perhaps we can make a move similar to Gray's but for the second condition instead of the first. For the sense of touch to count as a sense we may have to relax the sense-organ requirement somewhat, instead allowing for a variety of receptors in spread-out locations, as long as they are historically dedicated to facilitating behavior with respect to an identifiable physical class of energy. I will further explore this option in chapter 2.

1.6 Pragmatic leanings

The lesson I draw from the discussion of the four classical theories and the modern alternatives is that there is likely no right answer to the question what senses we have, if that question is seen as a metaphysical question that can only be answered with necessary truths.

The classical proposals had more metaphysical leanings, but they are all subject to serious objections. The contemporary proposals, I argue, all have a kind of pragmatism built in: the

---

[5] In chapter 2 I will say more about the polymodality of the transducers for thermal stimuli.

theories may give answers to what senses there are, but these are answers specific to an explanatory project, not Universal Metaphysical Necessary Truths.[6]

MacPhersons contemporary approach faces problems with providing criteria on how to select dimensions for the model, and how to scale and weight these dimensions. As it stands, the model is unlikely to give us any answers in the ontology of senses *per se*, but that doesn't mean such a dimensional analysis couldn't prove useful in more limited explanatory contexts. Plotting information about sensory systems in a multidimensional space and applying analysis to that space could provide answers to specific questions, especially related to quantifiable aspects of those senses. The selection of dimensions and the scaling and weighting could be informed by the type of questions one wants to answer through the analysis. This reflects scientific practice: models don't pop into being *ex nihilo,* they are constructed for a purpose.

O'Callaghan's account is overtly pragmatic – it provides an individuation schema which can be 'loaded' with criteria for individuating information extraction functions. Depending on how finely information extraction functions are individuated, the senses are also individuated more finely or more coarsely.

Keeley's account at the outset does not embrace much pragmatism about what sense modalities there are, but the criticisms force it in a more pragmatic direction. Nudds' criticism alleges that Keeley provides a principal suited for a scientific explanation, but not for the explanatory project of individuating the commonsense senses. Gray's counterexamples force the relaxation of the status of one of the criteria, but only given certain case-by-case considerations – which smells like pragmatism.

Fulkerson actively endorses a more pragmatic attitude towards individuating the senses, saying that the distinction between unimodal and multimodal depends on 'levels of explanation'. Saving Keeley's proposal from Fulkerson's criticism also leads to an adaptation in the direction of pragmatism. One of the requirements (2) is relaxed so as to be judged on a case-by-case basis, in the context of the other requirements.

Pending the invention of a better metaphysical theory of sense modalities (if it is possible at all), I will endorse a principled pragmatism towards the concept of a sense. Whether something counts as a single sense or not depends on the explanatory project one is engaged in. This thesis is an explanatory project in the philosophy of thermoception.

In the next chapter I will discuss the human thermoceptive system and discuss how we can see the sense of temperature as separate from touch.

---

6 One might say that finding Universal Metaphysical Necessary Truths is an explanatory project. What I mean however, is explanations that are limited in their scope of quantification and the information they take into account, and have defeasible conclusions.

2 Thermoception

In the previous chapter, I have discussed ways of individuating sense modalities. With the lessons learned from that discussion, I now turn towards the central topic of this thesis: the sense of temperature. It is an open question at this stage of the thesis whether humans have a sense of temperature. Depending on the principle of individuation for sense modalities one adopts, it may turn out that the ability to perceive warm and cold stimuli is not a sense proper, but rather part of another sense modality, i.e., touch.

The goal of this chapter is to introduce the reader to the science of temperature perception and use that science to argue a few philosophical conclusions. Before diving into the scientific literature, section 2.1 gets out of the way some terminological issues. Section 2.2 gives an overview of the neural mechanisms for transduction of thermal stimuli in the (human) skin. Section 2.3 is about thermoregulation: the ability of animals to regulate their body temperature. Section 2.4 follows the thermal signal up the afferents towards the central nervous system and summarizes how the peripheral signals are processed and contribute to perception and thermoregulation. In section 2.5 I integrate the lessons about sensory individuation from chapter 1 with the empirical summary presented in the preceding sections. I argue that the way temperature perception works, makes it a candidate to be counted as a proper sense modality on the neuroethological account. Specifically, I use section 2.5 to argue that thermoception is separate from the sense of touch. However, thermoception doesn't fit entirely neatly into the neuroethological account, so to accommodate this, I present a modified neuroethological account in section 2.6. In section 2.7 I discuss the status of thermoception as a part of the somatosensory system. In section 2.8 I discuss the 'dual modes' of thermoception, which were the subject of philosophical debate in the 1960s. In section 2.9 I discuss a contemporary distinction between haptic experience and tactile experience of touch. The discussion of this distinction sheds light on the old dual-modes discussion.

2.1 Terminology

The main subject of this thesis is the human ability to perceive thermal properties of the environment. That is a cumbersome phrase to use in every other sentence, so something shorter would be better. What term best captures this topic? What we need is not necessarily a term for a sense modality; later in this chapter I will argue that it makes sense to speak of a dedicated sense modality for thermal properties, so it makes sense to speak of a *thermal sense* or a *sense of temperature.* But sometimes I may want to talk about thermal perceptual abilities without presupposing that these abilities constitute a separate sense modality. And indeed, for many of the claims in this thesis it does not matter whether there is a dedicated sense modality for those perceptual abilities. So, for a general term for these perceptual abilities, I will avoid speaking of a *sense*.

Then, would '*temperature perception'* be the right term to use? To me, that phrase sounds more like the activity that a person (or animal) with thermal perceptual abilities might engage in, and not like a term for a faculty or ability. Moreover, as we will see later in the thesis, it is a subject of discussion whether our perception of thermal properties does track *temperatures* or some other thermal magnitude.

In psychology and physiology texts the terms '*temperature sensation'* or *'thermal sensation'* are sometimes used. Besides the reason mentioned above to avoid the word 'temperature', I also dislike using the word 'sensation' in this philosophical context because it carries the association of being about the experience of thermal properties, the phenomenology. In my investigation into the ability to perceive thermal properties I want to include the conscious experience of thermal properties, but I do not want to terminologically preclude myself from talking about the low-level neurophysiological aspects of the topic.

Another term frequently seen in physiology of psychology texts is '*thermoreception'*. This term has the advantage of not directly referring to temperatures as the thermal properties that are being tracked, but rather using the more neutral prefix '*thermo-'*. The downside of this term is that its close association with neurophysiology makes it seem that it only refers to the transduction of thermal stimuli into neural signals at the very periphery of the nervous system and doesn't include higher-level cognitive aspects and conscious experience. Just like I don't want the chosen terminology to emphasize too much the conscious/cognitive side of things, I don't want it to emphasize too much the low-level neurophysiology either.

The term I have settled on using is '*thermoception'*. Although it has been used in the literature, I introduce it here as something of a term of art. I take it to refer to the ability to perceive thermal properties of the environment, where 'perceive' should not be taken to imply conscious experience, and 'thermal properties' should not be taken to automatically mean temperatures. As such, it is an 'open' term, in that can refer to phenomenally rich conscious perceptual experiences that are perhaps paradigmatic in philosophy of perception, as well as unconscious sensory processes that some philosophers would not classify as perception, but that are important to organisms that coordinate their behavior and bodily processes to a thermal environment. I think of the term 'thermoception' as being on par with terms like 'olfaction', 'audition' or 'vision'. The term 'vision' includes conscious visual experience but also the retinal transduction of light into neural signals, just as I have stipulated that 'thermoception' can refer to both higher and lower-level processes. The term 'vision' also leaves open what properties of the environment count as visual properties. There is a good deal of research on what environmental features the visual system tracks or represents, and similarly for thermoception it can yet be determined what properties exactly the thermoceptive system responds to.

2.2 Temperature receptors

This section aims to give an overview of the neurophysiology and psychophysics of temperature perception. It is important in the context of this thesis to present some of this information, because it can help us answer philosophical questions about temperature perception. We can have no hope of answering the question what temperature experiences represent without accounting for the properties that the thermoceptive system is responsive to. Psychophysics and neurophysiology provide us information about the range of inputs a system is responsive to, but as I will argue later, the input conditions underdetermine what is represented by the system. The neurophysiological story is important to answering questions on mental representations, but it doesn't give a full account of the content of said representations.

This section contains no original research but is meant to introduce to an audience of philosophers some basics of how the thermoceptive system works. Most of the information in this section can be found in a review article by Schepers & Ringkamp (2010). I have added specific citations to claims that are based on specific experiments, but for the general neurophysiology of the receptors that has been well-established for decades I will simply refer to the review.

Thermoception is facilitated by four functionally distinct categories of primary afferent nerve fibers known as *thermoreceptors*. There are afferent fibers that react to cool stimuli (innocuous cold), warm stimuli (innocuous warmth), noxious cold, and noxious heat. These thermoceptive afferent nerve fibers have been well-known since the 1950s, largely due to the work of Herbert Hensel (e.g. Hensel, 1974, 1982, 1983; Hensel & Zotterman, 1951b) In the last 20 years or so transient receptor potential (TRP) ion channels have been identified as the molecular basis for thermal sensation. These TRP channels are sensitive to specific ranges of temperatures and are embedded in the terminals of afferent nerve fibers.

The receptors for innocuous warm temperatures and innocuous cool temperatures have differential static and dynamic response curves, meaning that they respond differently to temperature change than to steady temperature states.

Innocuous cold

The afferent nerve fibers that respond to cool stimuli exhibit ongoing action potential activity at normal skin temperatures.

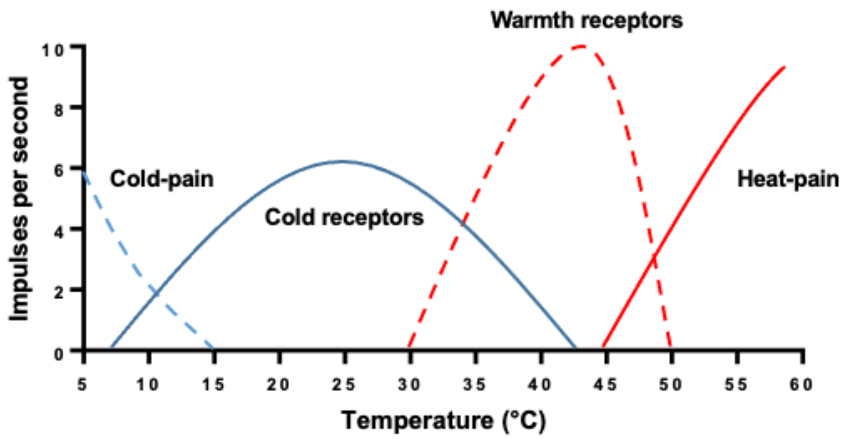


Figure 4: Receptor responses for static temperatures. Adapted from Guyton & Hall (2011).

At steady state temperatures the receptors for innocuous cold have a bell-shaped response curve, with maximal activity between 20°C and 30°C (see figure 4). Below and above this temperature activity decreases. At static temperatures below 17°C or above 40°C cold receptor discharge frequency becomes very low or zero. Some cold fibers are activated by high temperatures in the noxious range. This phenomenon may be at the basis of paradoxical cold sensation in response to hot stimuli.[1]

Cold receptors respond strongly to cooling of the skin and have inhibited response when the skin is warmed. This so-called dynamic response is transient: receptor response normalizes to the steady-state rate after a short period of time. Cold receptors can be chemically activated, especially by menthol. At cool temperatures (when the receptors are active) applying menthol increases receptor discharge rate. Menthol *sensitizes* cold receptors. The application of menthol to the skin correspondingly elicits a cool sensation. Cold fibers do not respond to mechanical stimulation. Cold fibers are thinly myelinated (Aδ) afferents, as can be determined from their conduction velocity. (Darian-Smith, Johnson, & Dykes, 1973; Dubner, Sumino, & Wood, 1975; D R Kenshalo & Duclaux, 1977)

Other nerve fibers exhibit response to stimuli in the innocuous cold range, but their contribution to cold perception is smaller or negligible. Among the other fibers that respond to innocuous cold are 1) fibers that respond to cold stimuli and are unresponsive to mechanical stimulation. These are known as High threshold Cold Receptors (HCR). HCRs respond to temperatures in the low end of the innocuous cold range and adapt quickly to 'neutral' temperatures between 20°C and 30°C, meaning that they do not have a steady response at maintained temperatures. HCRs do not respond to hot stimuli or mechanical stimulation. HCRs have conductivity velocities in the low Aδ or C fiber range. It has been suggested these HCRs contribute to innocuous cold sensation at temperatures below the maximal response of 'normal' low-threshold cold receptors. (LaMotte & Thalhammer, 1982)

Another fiber type that's responsive to cool stimuli are 2) Unmyelinated low-threshold mechanoreceptors (CLTM) that respond to rapid cooling. CLTMs do not respond to static temperatures, but only to rapid cooling. They respond strongly to touch, and not at all to heat. Since the response to cooling is small compared to the response to touch, it is doubtful CLTMs contribute to thermoception. (Kumazawa & Perl, 1977) Experiments with differential nerve blocks have shown that colds sensation is mediated by Aδ fibers, and not C fibers.

Among mechanoreceptors we also see sensitivity to cool stimuli; 3) the large myelinated fibers of slowly adapting mechanoreceptors (SA fibers) sometimes respond to cool stimuli.[2] Ruffini endings in deeper skin layers and Merkel discs in superficial skin layers respond to below normal skin temperature and above 14,5°C. However, the response to temperature is limited in comparison to the response to mechanical stimulation. As a result, it is unlikely that SA fibers contribute much to thermoception. (Johnson, Darian-Smith, & LaMotte, 1973)

---

[1] I characterize this paradoxical sensation as a thermoceptive illusion in chapter 6.

[2] The response of mechanoreceptors to cooling may be the explanation of a thermoceptive illusion known as the silver Thaler illusion. I will say more about thermoceptive illusions in chapter 6.

The principal transducer molecule for cold sensation is the transient receptor potential melastanin 8 (TRPM8), an ion channel activated by temperatures below 26°C. TRPM8 is also activated by chemical stimulants known to cause cold sensation, such as menthol and ilicin. In animals that have been genetically manipulated through a gene knock out technique not to have TRPM8 have less sensitivity to innocuous cold. However, response to noxious cold is much less affected by abolishing TRPM8. While response to innocuous cold is reduced, it is not absent, which shows there are other molecular mechanisms for the transduction of innocuous cold. (Bautista et al., 2007) TRPM8 and other cold-transducing mechanisms can be present at the same terminals.

Innocuous warmth

Warm fibers are continuously active at temperatures above 30°C and are not responsive to mechanical stimuli. Like cold fibers, the steady-state response curve is bell-shaped, with maximal response around 40-43°C and minimal discharge below 30°C or above 50°C.

Also, just like cold fibers, warm fibers respond to dynamic temperature: the response of warm fibers is strongly inhibited by cooling and warming results in a burst of activity that returns to steady state levels. Warm fibers are C fibers and warmth sensation disappears when C fibers are blocked. (Mackenzie, Burke, Skuse, & Lethlean, 1975)

The transient receptor potential vanilloid 3 (TRPV3) ion channel is activated by temperatures above 33°C. TRPV3 is activated and sensitized by camphor, which is known to elicit a sensation of warmth. This protein is expressed in other tissues besides afferent nerve fibers such as in the dorsal root ganglion and in epithelial cells in the skin, on the tongue and in testes. TRPV3 on the tongue responds to chemical stimuli (cloves, oregano, thyme) and may be involved in taste perception.

Mice in which the gene for TRPV3 has been knocked out show less ability to coordinate their behavior to temperatures in the environment and also have diminished behavioral response to noxious temperatures. Warmth sensation is not entirely absent in these knock out mice, which indicates other mechanisms contribute to warmth sensation.

Another ion channel also responds to warm stimuli. TRPV4 is activated by temperatures above 24°C or 34°C. (Schepers & Ringkamp, 2010). The ion channel is also activated by hypotonic stimuli and its function may be a stretch-receptor or osmosis sensor. Temperature behavior in mice where the gene for TRPV4 is knocked out changes: the mice gain a preference for higher innocuous temperatures. Whether TRPV4 plays a role in noxious heat sensation is unclear.

Noxious cold

Thermoreceptors for noxious cold are inactive at normal skin temperatures. The temperature at which painful cold sensations are felt varies across the body: about 10-15°C in glabrous skin, and 18°C in hairy skin. Slow cooling of the skin is less painful than rapid cooling; just like with innocuous temperatures there is a different response to static vs. dynamic temperature. The intensity of pain is linearly correlated with temperature from 0-20°C with lower temperatures being more painful.

Noxious cold stimuli can cause different kinds of pain sensations such as pricking, burning/heat, and aching. A variety of afferent fibers respond to noxious cold stimuli, which may explain the different experiences of cold pain.

Both Aδ and C fibers are involved in the sensation of noxious cold. Threshold values vary between fiber populations, but the average response and peak discharge of both cold-responsive Aδ and cold-responsive C fibers increases as temperature decreases. Both the Aδ and C fibers involved in cold sensation are polymodal, in that they also react to mechanical stimuli.

At temperatures of 2-3°C the unmyelinated fibers in superficial nerve endings become less conductive, yet cold pain persists at those temperatures and below. This likely indicates that deeper (peri)vascular nociceptors contribute to the sensation of very cold temperatures.

When Aδ fibers are blocked, the pricking sensation of cold disappears. In conditions where Aδ fibers are blocked but C fibers remain active, noxious cold is felt as heat or a burning sensation. This indicates that the signal from Aδ fibers modifies or blocks the C fiber signal to the central nervous system. Sensation of noxious cold temperatures depends on this integration of signals from various fiber types.

Transduction of noxious cold possibly involves TRPA1, as this ion channel responds to temperatures of 17°C, which is lower than the response temperature for TRPM8, which is involved in sensation of innocuous cold. However, research with gene knockout mice and in-vitro electrophysiology of TRPA1 is inconclusive.

Noxious heat

A variety of fibers are involved in noxious heat sensation. In the sensation of noxious heat on hairy skin we can usually distinguish two distinct pains: a fast and sharp pain and a slower dull or burning pain that follows about 1-2 seconds later. The fast first pain is mediated by Aδ fibers (that have higher conduction speed) while the delayed second pain is mediated by slower conducting C-fibers. The fast, sharp pain disappears when an A-fiber block is applied.

The Aδ fibers involved in noxious heat sensation can be divided into two categories: Type I afferents that have a high threshold (above 53°C) for short heat stimuli, and a delayed response (5s) to longer (30s) heat stimuli. Burn injuries can sensitize these fibers, which explains thermal hyperalgesia in these cases.

Type II fibers have a lower threshold (around 47°C) and respond quickly (1s) and strongly to shorter stimuli. These afferents adapt when the stimulus continues. Type II fibers are not found in glabrous skin and as a consequence there is no first pain sensation to stimuli applied to e.g., the palm of the hand.

The C-fibers involved in noxious heat sensation have thresholds for temperature response from 37°C to 49°C. Normally the pain threshold for heat is around 45°C. The C-fibers contribute to heat pain sensation through temporal summation in the CNS.

Most C-fibers that are responsive to heat are also responsive to mechanical stimuli, and only a smaller portion of C-fibers are unimodal heat receptors.

Lastly, a subset of afferent fibers that are responsive to innocuous warm stimuli are also responsive to noxious heat. These fibers have the static response function of warm fibers, but the peak discharge rate lies at a noxious temperature (45°C).

The transducer molecules responsible for heat sensation are TRPV-1, TRPV-2, and TRPV-3. TRPV-1 is activated by temperatures above 43°C. TRPV-2 is activated by temperatures above 52°C and TRPV-3 is activated by temperatures above 50°C. TRPV-1 knockout mice still show normal behavioral responses to temperatures in the noxious range, so other transduction mechanisms besides TRPV-1 must function in the range from 43°C to 50°C. TRPV-1 is activated by capsaicin (the compound that makes peppers spicy) and low pH. TRPV-2 and TRPV-3 are not sensitive to low pH or capsaicin.

2.3 Thermoregulation

Humans (and other mammals) have to maintain a body temperature within a rather narrow range to survive. The ability to maintain a core-body temperature within such a narrow range even as environmental temperatures vary greatly is known as Homeothermy. When core temperatures drop too low this leads to a pathophysiological state known as hypothermia whereas abnormally high core temperatures lead to hyperthermia. Both these pathological states can result in death. For humans, core temperature is normally maintained within a range of 36.5°C to 37.5°C. (Tansey & Johnson, 2015; Terrien, Perret, & Aujard, 2011)

Humans and other mammals are endotherms, meaning that internal metabolic processes provide most of our body heat. Maintaining a stable core temperature is done by maintaining a balance between heat production and heat loss. Thermoregulatory processes can be divided into two categories: autonomic thermoregulation and behavioral thermoregulation. Within both these categories we can distinguish strategies for avoiding hypothermia and for avoiding hyperthermia.

Autonomic mechanisms for thermoregulation are largely unconsciously controlled and executed, although we may feel cold or hot and can be aware of the effects of autonomic thermoregulatory processes. The autonomic thermoregulatory system is driven by inputs from thermoreceptors in the skin, the hypothalamus, the spinal cord, the great veins, and viscera, and controls processes that influence thermogenesis (heat production) and thermolysis (heat dissipation). (Tansey & Johnson, 2015)

To avoid hypothermia, autonomic processes can either increase heat production (thermogenesis) or decrease heat dissipation (thermolysis). Thermogenic processes that can be triggered to avoid hypothermia are: non-shivering thermogenesis in brown adipose tissue, shivering in skeletal muscles, and an increase in metabolic rate. Autonomic processes that diminish thermolysis in order to avoid hypothermia are: vasoconstriction in the blood vessels of the skin and piloerection (hair raising). (Tansey & Johnson, 2015)

To avoid hyperthermia, the thermoregulatory system engages opposite responses: it either decreases heat production or it increases heat dissipation. To decrease thermogenesis the metabolic rate may be lowered through the adrenal and thyroid glands. Autonomic processes that increase thermolysis in order to avoid hyperthermia are vasodilation in the blood vessels of the skin, and sweating. (Tansey & Johnson, 2015)

Thermoregulation is closely tied up with the management of energy expenditure. Most of the heat produced by mammals comes from muscle activity, which demands a lot of energy.

To save energy on processes of autonomic thermogenesis (such as shivering), mammals engage in behavioral thermoregulation that influences thermogenesis or thermolysis.

Increased activity is a behavioral way of avoiding hypothermia, as increased muscle activity leads to more heat production. Increase in energy intake (through eating) is also a behavioral way of avoiding hypothermia: more available energy means that energy-consuming thermogenic processes can be maintained. Other behavioral strategies for avoiding hypothermia aim at minimizing thermolysis. Such behaviors include adopting a ball-like posture, basking, huddling, nest-building, nest-housing[3], and nest-sharing. (Terrien et al., 2011)

To avoid hyperthermia animals may decrease their activity (to minimize muscular heat production) and decrease energy intake. Habitat selection can help increase thermolysis (e.g., moving to shaded areas or burrowing), as can adopting certain postures that maximize body heat loss. Panting can also aid thermolysis, for example in dogs.

Humans have an especially varied repertoire of thermoregulatory behaviors. For example, constructing buildings with features that allow us to cool or warm our environment is essentially a sophisticated version of the burrowing or nest-building other animals engage in. In a later chapter I will say more about the advanced behavioral thermoregulation of humans and what that means for our perceptual capacities.

2.4 Temperature sensation in the central nervous system

Information from thermoreceptors is processed in the central nervous system in a variety of ways along a variety of pathways.

Thermal receptors in the central nervous system

Besides the thermoreceptors in the skin, mammals have thermoreceptors in the spinal cord, the viscera and in the brain. These function as inputs for the thermoregulatory system, providing information about core temperature. Thermoreceptors in the skin also function as input to the thermoregulatory system: for a timely response that avoids hyperthermia or hypothermia it is useful to react to peripheral temperature changes, and not only to core temperature changes. (Tansey & Johnson, 2015)

Lateral parabrachial nucleus pathway

Afferent signals from peripheral innocuous warm and cool receptors travel through separate primary somatosensory neurons in the dorsal horn to the lateral parabrachial nucleus (LPB), from where neurons project to the median preoptic nucleus (MnPO) in the preoptic area (POA) of the hypothalamus. From the POA mechanisms of heat gain or heat loss are controlled, specifically those leading to metabolic changes in brown adipose tissue, vasoconstriction or vasodilation in blood vessels in the skin, and shivering thermogenesis in skeletal muscles.

---

[3] Nest-housing: utilizing an existing shelter as a nest rather than building it.

It was previously thought that the POA functioned something like a thermostat, integrating signals from peripheral and central thermoreceptors, and comparing the integrated signal to some set point. Deviations from the set point were thought to trigger thermoeffector responses. New techniques have revealed this model is likely false. The current view of central thermoregulation is that it does not use a set point and integrated signal, but rather that individual thermoeffector circuits can be triggered by peripheral or central temperatures. (Nakamura, 2011; Romanovsky, 2007; Tansey & Johnson, 2015)

Spinothalamocortical pathway

A very interesting recent development in the neuroscience of thermoception is the discovery that the pathway that drives autonomic thermoregulation in rats is different from the pathway that underlies the ability to discriminate warm and cool temperatures. Conscious temperature sensation is realized by the spinothalamocortical pathway. The temperature signal travels from peripheral neurons via the dorsal horn to thalamus, from where it projects to the primary somatosensory cortex.

In studies with rats, it has been shown that even behavioral thermoregulation is the result of activation of the LPB pathway rather than the spinothalamocortical pathway. (Yahiro, Kataoka, Nakamura, & Nakamura, 2017) In a later chapter I will say more about what this discovery means for the relation between conscious thermoception and thermoregulation.

Nociceptive pathway

Thermal pain signals from thermosensitive nociceptors are processed differently from innocuous temperature stimuli. As mentioned in section 2.2 most thermosensitive nociceptors are polymodal. This is generally true for nociceptors.

For a long time, theorizing about pain has been dominated by the thought that pain is something like a sense modality that allows us to perceive (impending) damage to the body. The so-called *specificity theory of pain,* based on the work of Müller (1835) and von Frey (1895) held that there are elements of the central and peripheral nervous system specifically devoted to the processing of painful stimuli. The competing *pattern theory of pain* held that pain is realized by certain spatial and temporal patterns of activation, regardless of modality. The advent of new research techniques has largely settled the matter in favor of the specificity theory, although certain elements of the pattern theory still play an important role: there is considerable plasticity in nociceptive channels, and the function of peripheral nociceptors and nociceptive central neurons can change dramatically, especially in damaged tissues. (Cervero, 2008)

Pain signals from first order neurons located at the dorsal root ganglia are transferred to second order neurons in the posterior horns of the spinal medulla that ascend through the spinothalamic tract. Third order neurons project to the somatosensory cortex and the cingulate cortex. Many structures are involved in the complex process that is pain. The parabrachial nucleus, intralaminar nucleus of the thalamus, and the amygdala are involved in cognitive and affective dimensions of pain. (Aguggia, 2003)

2.5 Thermoception and the sense of touch

Is our sense of temperature separable from the sense of touch? Historically thermoception has sometimes been classified as a part of touch, for example on the Aristotelian five-senses view. In textbooks related to the senses thermoception is often grouped in a chapter or volume with touch, although not necessarily seen as a part of it. On the other hand, philosophers, biologists, and psychologists have no trouble researching thermoceptive abilities separately from touch and vice versa.

Matthew Fulkerson has argued for a theory that sees touch as a unified sense based on *feature binding*. (Fulkerson, 2011, 2014) The idea is that in tactile experience, various tactile features are bound to the same perceptual object: for example, when you hold a squash ball, its round shape, its weight and its grippy texture are experienced as being qualities of the same object. That is not to say that you can't perceive texture without perceiving weight. But according to Fulkerson, fundamentally unisensory experiences attribute properties to one perceptual object. In the case of touch, this can include thermal properties. If the squash ball you are holding was just used in a game of squash, it will be warm to the touch, in addition to the properties mentioned above. For Fulkerson, the binding of texture, temperature, shape, and weight to this single object makes it a unisensory experience. This means that thermal perception can be counted (in this case) as a part of touch.

While I think Fulkerson gives an accurate description of the phenomenological fact of feature binding, I don't think this constitutes an argument that thermoception *is* touch.[4] In many cases of thermoception, the temperature experience isn't bound to an external object at all, but rather to our own body. Add cases of radiant heat perception (e.g., turning your face to the warmth of the sun) that aren't bound to tactile experience, and it starts to look like feature binding of tactile and thermal stimuli isn't ubiquitous, but rather a special case.

Mohan Matthen has defended a view of sensory individuation that distinguishes between *sensory* modalities and *perceptual* modalities. (Matthen, 2015) Sensory modalities are individuated by receptor types, while perceptual modalities are individuated by the activity one engages in when employing the modality. On the sensory level, Matthen acknowledges that haptic touch and thermoception are distinct. On the perceptual level, however, he claims that the experiences that follow from *touching* belong to the perceptual modality of touch. The experience of the temperature of a cup of coffee you are holding is the product of *touching*, and therefore belongs to the perceptual modality of touch. Cases in which you feel a temperature without touching an object (as with radiant heat from the sun) are categorized differently by Matthen, as we will see later in this chapter when we discuss his tactile/haptic distinction. (Matthen, 2021)

Ratcliffe (2012) has argued that the sense of touch (which he takes to include thermoception) cannot be principally individuated from other senses, but neither can it be divided up into separate senses. Although Ratcliffe does not make this explicit, his position seems to imply that there can be no separate sense of temperature perception.

---

4 Note that Fulkerson doesn't make that claim either.

His argument for the claim that touch cannot be individuated is that there are simply no characteristics (phenomenological or non-phenomenological) that are common to all perceptual activities that we count as touch which exclude other senses. He argues (as does Fulkerson) that the four Gricean criteria as well fail to individuate touch and he aligns with Gray (2005) in a rejection of Keeley's account of sensory individuation. I give arguments for a modified version of Keeley's account below. The claim that touch cannot be divided up into separate senses is argued for by appeal to phenomenology: in some touch experiences sensations of temperature and pressure combine to form experiences in which those elements are not clearly phenomenally distinct. I find that argument unconvincing: the existence of multimodal experience is not an argument against the existence of separate sense modalities.

Furthermore, Ratcliffe appeals to the polymodality of receptors involved in touch to argue that even on a neurophysiological level there is no sharp distinction between a pressure sense and a temperature sense. I disagree with this interpretation of the empirical literature: while there is certainly polymodality, its influence on the perception of temperature is limited. I view the perceptual effects of polymodal receptors as a special case, not a central case. I will return to these effects in chapter 6.

The four criteria from Keeley's account of sensory individuation can help us to see in what ways thermoception differs from or aligns with the sense of touch.

(1) Physical stimulus

In terms of physical stimulus thermoception seems *prima facie* to be different from other forms of touch. Haptic touch depends on the skin (or hair) being in contact with objects or substances that mechanically stimulate the skin. Thermoception responds to temperature and changes in temperature.

Neurophysiological literature usually refers to the cutaneous receptors involved in touch as *mechanoreceptors,* while the receptors involved in thermoception are referred to as *thermoreceptors*. The fact that thermoreceptors can be polymodal both at the neural and at the molecular level does not imply that thermal stimuli can't be differentiated from mechanical stimuli, and in fact such confusions are rare.[5]

But recall the argument about the thermosensitive nose of vampire bats made by Gray. The argument was that the thermoreceptors in that case respond to not one but two types of physical stimuli: radiant heat and conducted heat. Conducted heat is a type of kinetic energy, as is mechanical stimulation. Does that mean that thermoception and touch partially overlap in what physical stimuli they are responsive to? That depends on how you individuate physical stimulus types. Gray's argument about vampire bats relies on cutting up the types of physical stimuli in a certain way, that I think is not productive.

One principled way of individuating physical stimuli is to look at physics for what types of energy there are. The most fundamental types of energy that physics recognizes are electromagnetism, gravity, strong interaction, and weak interaction. In principle, the

[5] An example of such confusion is the silver Thaler illusion, where thermoceptive stimuli influence the perception of weight of an object on the skin. I will discuss this illusion in chapter 6.

mechanical impact of an object on a mechanoreceptor is explainable in terms of these four forces. The concept of 'radiant heat' that Gray appeals to is not a fundamental physical category: it is just a name for electromagnetic radiation in a certain part of the spectrum. Kinetic energy is, arguably, also not a fundamental type of physical energy. Gray doesn't rely on the fundamentality of the physical stimulus types – he forgoes that option of providing a principled way of individuating stimulus types. Gray allows that in certain cases (such as the pit viper case) we may want to count separate parts of the electromagnetic spectrum as separate stimulus types. I would like to take this pragmatism a step further: in specifying stimulus types we do not have to appeal to fundamental energy types, we only must be able to describe and differentiate the stimulus type in physical terms.

To take the example of the pit vipers: the different ranges of the electromagnetic spectrum that the eyes and the pits are sensitive to can perfectly be described and differentiated in physical terms. Both stimulus types are described as being electromagnetic radiation, and they are differentiated in terms of the frequency of the radiation. That physics doesn't see different ranges of the electromagnetic spectrum as different energy types doesn't mean that biology can't take them to be different stimulus types, especially when they are so neatly defined in physical terms. The same goes for the difference between the kinetic stimulation that mechanoreceptors are sensitive to and the conductive heat that thermoreceptors are sensitive to. Conductive heat and mechanical stimulation may be the same physical energy type (kinetic), depending on how you explicate that notion, but that doesn't mean they are the same stimulus type. Now, conductive heat and mechanical stimulation are not different ranges of the same spectrum as in the pit viper example, but rather they are different physical processes best described by different branches of physics. So, while we may not have reasons grounded in *fundamental* physics to distinguish the two, we certainly have the tools in physics to describe the processes these receptors are sensitive to and differentiate between the processes.

What physical stimulus type then is thermoception responsive to? My claim is that thermoception represents temperature and changes in temperature. This claim needs defending, and I will do so in later chapters.

(2) Sense organ

As explained above in the sections on the neurophysiology of thermoception, there is not a single dedicated sense organ for thermoception. Rather there is a variety of functionally and physiologically distinct receptors, that are often polymodal. On both the neural and molecular level there is differentiation and polymodality: a range of temperatures is serviced by multiple receptor types, and receptors may also react to stimuli in other modalities. Because of this, thermoception fails the requirement of having a single dedicated sense organ.

Recall a similar argument we encountered in the first chapter: Fulkerson argued that Keeley's account could not count touch as a single and distinct sense modality because of the variety and distribution of the receptors for touch. There isn't a single dedicated sense organ for haptic touch. Fulkerson's response is to abandon Keeley's strategy of individuation, but I offered the suggestion that the sense organ criterion (2) could perhaps be relaxed. Just like we could adapt Keeley's view to count the sense of touch as a sense modality by relaxing (2), we can do the same for thermoception. While the sense organ for temperature is not as easily recognizable as an eye or the ear, there are physiological structures dedicated to perceiving temperatures, and we should count those as the sense organ(s) of thermoception.

Although thermoception does not have a single dedicated sense organ, it does have an associated class of receptors that function as thermoreceptors. There is not a single organ of the sense of temperature, but we do have a good grasp of the physiology of the different kinds of receptors involved in thermoception. Besides, we have a good idea on how the stimulation of these receptors contribute to *experiences* of hot and cold, as described by psychophysics. Furthermore, we have an idea of how the signaling of temperature in the skin contributes to thermoregulation. We should all be familiar with behavioral responses to temperature – we routinely adapt our behavior to the temperature of the environment, and we see other people and animals do the same.

Together, this doesn't exactly form a picture where biologists are at a loss as to what could be the sense organ of temperature. Sure, there is work that remains to be done in identifying the molecular receptors that underly the thermoceptive function of afferent fibers and in identifying the neural mechanisms of thermoception and thermoregulation. But overall, we have a good picture of how human bodies receive temperature information from the environment. It seems foolish to maintain on philosophical grounds that senses can only have a single organ if a scientific account of a sense involving multiple receptors is available.

It is my claim therefore that we should adapt (2) to allow for multiple sense organs for a single sense modality, if we can show that these sense organs form a natural group in terms of the physical stimuli they respond to, the role they play in the ability to behaviorally discriminate stimuli, and their dedication to salient parts of the organism's environment. In other words: criteria (1), (3), and (4) can help us decide how we can relax requirement (2) to include multiple sense organs under the same sense modality.

A problem that remains for the idea that thermoception has multiple dedicated sense organs is the fact that thermoreceptors can be responsive to chemical or mechanical stimuli. However, the part that such responsiveness plays in our behavior is limited. Cases where non-temperature stimuli are perceived *as temperatures* should be explained as cross modal illusions, I will talk about this in a later chapter.

(3) Behavior

The third requirement on Keeley's account was that a sense should enable us to behaviorally distinguish between stimuli within the stimulus type. We can discriminate between stimuli of different temperatures based on their temperature without appealing to haptic aspects of the object. Hence the third requirement helps us distinguish thermoception from touch.

(4) Dedication

The fourth criterion on Keeley's account was that the putative sense should be of evolutionary or developmental importance to the organism. This is clearly true for perception of temperature. The ability to coordinate our behavior to environmental temperatures plays an important role in our lives, as behavior is an important aspect of thermoregulation. Thermoregulation is essential to the survival of homeothermic animals, and behavioral thermoregulation plays an important role in this. As explained in section 2.3, behavioral thermoregulation allows organisms to make large changes in the balance between heat production and heat loss and reduces the energy expenditure of autonomous thermoregulation.

To conclude this section: we can use Keeley's four criteria to distinguish touch from thermoception if we modify (1) and (2). There may be arguments to view thermoception as a part of touch in some explanatory projects, but I take the discussion above to show there is nothing wrong with thinking of thermoception as a sense modality distinct from touch.

2.6 Modified neuroethological account of the senses

In dealing with counterexamples to Keeley's account, I have made some modifications to the criteria and the ways these criteria combine to provide principles of sensory individuation. In this section I will present the product of those improvements as an alternative to Keeley's account. I call this the *modified neuroethological account of the senses.*

The driving idea behind the modified neuroethological account is to allow for a wider range of sense modalities. Specifically, the idea is to be able to distinguish sense modalities that may be responsive to the same physical quantity (such as pit viper vision and pit viper heat sense) and allow sense modalities that involve more complicated receptor setups, such as touch or thermoception.

The modified neuroethological account features the same four criteria as Keeley's neuroethological account, but it redefines the first and second criteria in less strict terms. Just like on Keeley's account, the four principles are individually necessary and jointly sufficient criteria for sense modalities. However, there is a strong internal relation between the criteria. Criteria (1), (3) and (4) inform how criterion (2) is to be applied, and criteria (2)-(4) provide a backdrop against which a plausible physical description of the stimulus type (criterion 1) should be formed. The modified neuroethological account that motivates awarding thermoception the status of a sense modality looks as follows:

> Stimulus type: The external physical stimuli to which the sense responds. Senses are distinguished by sensitivity to physical properties of the environment of the organism. Stimulus types are physical quantities or ranges of quantities or properties or sets of properties that allow for a unified physical characterization. Stimulus types should be individuated by appeal to the following criteria:
>
> 1.1 The stimulus type is describable in physical terms as a unified phenomenon. This excludes senses that respond to wildly different physical stimuli: there is no single sense for gravity *and* electromagnetic radiation.

1.2 The putative sense involves transducers for the putative stimulus type. If there is no transduction of that physical stimulus, it cannot be the stimulus associated with the sense modality.
1.3 The stimulus type associated with a particular sense is limited to that range of properties, which can be behaviorally discriminated by the organism that possesses the sense. For example, infrared light is not part of the stimulus type for human vision, because vision does not allow people to behaviorally distinguish stimuli in that range of the electromagnetic spectrum.
1.4 The property or sets of properties that define the stimulus type should be of evolutionary or developmental importance to the organism. This need not hold for the entirety of the range of stimuli an organism is responsive to. For example, we may be able to discriminate some smells that are not of direct evolutionary or developmental importance. Such a trait can be an evolutionary 'spandrel' – a trait with no adaptive advantage – that can be a byproduct of the evolution of some other trait. (Gould & Lewontin, 1979)

Sensory physiology: Senses are distinguished by the physiology and neurology involved in the transduction, encoding and processing of the physical stimuli. Senses need not have a single dedicated sense organ: the task of gathering information about a specific stimulus type may be distributed over various physiological structures, and structures involved in sensing stimuli in one modality may combine that task with other functions.

2.1 There must be transducers for the putative stimulus type, that are appropriately wired up to transfer information about that physical stimulus to the central nervous system. These transducers need not be exclusively dedicated.
2.2 The sensory physiology may involve different structures if they contribute to the organism's ability to behaviorally discriminate within a certain stimulus type.
2.3 The sensory physiology may involve different structures if these structures are evolutionarily or developmentally dedicated to detecting a physical stimulus type that is a part of the normal environment of the organism.

Criteria (3) and (4) are essentially unmodified from Keeley's account.

2.7 Thermoception and the somatosensory system

Thermoception is usually classified as part of the somatosensory system, along with touch, hunger, pain, and proprioception. The somatosensory system is usually defined as the sensory systems dedicated to tracking changes within the body or at the surface of the body. Related distinctions include the distinction between the special senses and the general senses, and interoception/exteroception. The question in this section is whether belonging to one of these categories precludes a sensory system from being a sense modality proper. I will argue that it does not.

Special/general senses

In medicine, a distinction is commonly made between the *special senses,* such as vision and audition, that have dedicated sense organs, and the *general senses* such as touch and thermoception that have receptors distributed throughout the body. Above I have argued that the fact that thermoception uses different types of receptors spread throughout the body should not be taken to mean that it is not a sense modality. On the modified neuroethological account the general can equally well be sense modalities proper as the special senses can. I don't think there is anything particularly wrong with making a distinction between those senses that have a more unified sense organ (e.g., an eye) and those that don't, but it is of little importance to sensory individuation.

Sensory / Somatosensory

The somatosensory system as usually defined includes touch, thermoception, hunger, pain, and proprioception. Some of these sub-systems of the somatosensory system (particularly touch) are commonly thought of as sense modalities, while others are usually not. On the modified neuroethological account, classification of touch and thermoception as parts of the somatosensory system does not influence their status as sense modalities.

Interoception / exteroception

One distinction related to the somatosensory system is that of interoception vs exteroception. Interoception is perception (or something like it) of internal states of the body. Hunger is a paradigm example: it reports a state of the body through internal mechanisms. Other parts of the somatosensory system are less clearly interoceptive: touch (especially haptic touch) can be used to gather information about the environment and not just about the body itself. It can only do so, however, for parts of the environment that touch the body.

The distinction between interoception and exteroception does not coincide with the distinction between the somatosensory system and the special senses. Hunger and proprioception are perhaps purely interoceptive, but touch is clearly exteroceptive to a certain extent.

Thermoception is a mixed case. Thermoregulation depends on both the monitoring of internal body temperature and of surface temperatures, i.e., temperatures in the environment. However, discriminating between temperatures of external objects is also a thermoceptive ability that humans possess, and is clearly exteroceptive. As such, we can distinguish between an interoceptive component and an exteroceptive component to thermoception.

2.8 The dual modes of temperature perception

The proximal or distal attribution of felt temperatures was the subject of a philosophical debate in the 1960's, most notably between G.N.A. Vesey and D.M. Armstrong. The debate traces its historical roots back to Berkeley, who aimed to show that perceived qualities (heat, taste, color, etc.) are mind-dependent; they cannot exist without the mind of the observer. Berkeley argues that the perception of great heat is *pain*, and the perception of warmth is *pleasure*. Since pain and pleasure cannot exist without the mind, perceiving heat is not perceiving a mind-independent property. The pain we feel when touching something hot is not attributed to the object, but to the body itself – we do not perceive something external. (Berkeley, 1713)

In a 1951 paper Gilbert Ryle notices something related, albeit from a linguistic perspective: the verb 'to feel' may be used in a variety of ways, one amongst which is the *perceptual use* – the way we use the verb when we say we feel the sweater to be soft, the box to be heavy, the bowl to be hot, etc. But there is a way of employing this perceptual use to describe bodily sensations that is of interest to this debate:

> […] while I might find out that my feet were cold with my hand or with a thermometer, ordinarily I find it out without employing either instruments or other bodily organs. Should I say that I feel that they are cold in or with my feet themselves? In real life I do not say this. I just say that I feel they are cold. (Ryle, 1951, p. 193)

Within the perceptual use of 'to feel' there are distinct situations in which you use parts of your body to feel external objects, or you just feel certain parts of your body to be in a certain state themselves.

The participants of the later debate were not particularly convinced by the idealistic conclusion Berkeley draws, nor very interested in analyzing the subtly different usages of the verb 'to feel'. What Vesey and Armstrong picked up on was the observation that sometimes sensations of touch or temperature are attributed to the external object that causes them, while at other times they are attributed to the part of the body that receives them.

> It is possible to distinguish between a person's perception of the heat of something and his feeling hot himself. Standing on a station platform in icy weather, one's feet are likely to feel cold. But their feeling cold is not a matter of their feeling cold to one's hand, although they might well do so. With one's hand one can feel the coldness of one's feet in the same way as, on moving one's feet from a warm part of a bed to a cool part, one can, with one's feet, feel the coldness of the sheets. But again, feeling the coldness of the sheets is not a matter of one's feet feeling cold, although, if the sheets are very cold, one's feet may soon come to feel cold. (Vesey, 1960, p. 202)

The debate between Armstrong and Vesey centers on the question of how to explain these dual modes of the bodily senses, and in particular thermal sensation.

Both Armstrong and Vesey acknowledge an initial distinction between two classes of bodily sensations. One comprises of sensations such as pains, tickles, and itches, the other of thermal sensations, pressure, motion etc. These classes differ in that the first class (of pains and tickles etc.) cannot represent those properties as being outside the body: pains are felt at a certain location in the body, not at the object that causes it. The second class, however, can manifest in two ways: as the dual modes recognized by Berkeley and Ryle, and explained above by Vesey.

Vesey calls the experience of the object *perceiving sensation* and the experience of one's own body *intrinsic sensation*. Itches and pains are always intrinsic sensations, and do not give rise to perceiving sensations. Up to this point Armstrong and Vesey agree. It is the explanation of the senses that can give rise to both modes of sensation that is the topic of controversy.

According to Vesey, all bodily sensation are essentially intrinsic sensations. When you feel your foot to be cold without the use of another body part, what you are really experiencing is an intrinsic sensation that *does not* attribute any temperature to any part of the body, because intrinsic sensations are unlike perception in that they do not attribute properties to objects.

The idea is that when you feel your foot to be hot, the content of that sensation is dependent on a previous perceiving sensation. At some earlier stage, you have felt your foot to be hot using some other part of the body, while at the same time having a particular intrinsic sensation. At that point an association is formed between this specific intrinsic sensation and the perceiving sensation of hotness. At later points in time, you may now recognize the intrinsic sensation as one of hotness, because of the association previously made. (Vesey, 1960)

The dual modes of temperature perception then are explained by the activity of two distinct processes, one perceptual and one not. The non-perceptual process of *intrinsic sensation* acquires its 'hotness' content through learned association with the perceptual process of *perceiving sensation*.

A competing explanation of the two modes of temperature experience is given by Armstrong (1962, 1963). According to Armstrong bodily sensations can be perceptual, and this is the case with heat sensations. He argues that what he calls the *tactile perception* of temperature (feeling your foot to be cold with your hand) and *bodily perception* (feeling your foot to be cold without using your hand) are distinct, but both are forms of heat perception.

On Armstrong's view, contrary to Vesey's, bodily sensation is a form of perception, and attributes properties to parts of the body. When you feel your foot to be cold, that ascribes a property to that part of the body. Consequently, bodily sensation of temperature can be erroneous. We may experience our feet to be hot when they are in fact cold. Vesey on the other hand thinks that bodily sensations of temperature are like itches: that some part of the body feels a certain way is a fact of experience and is not defeasible. That you have a tickle at a particular location is not defeated by the observation there is nothing tickling you. For Vesey, this is true of all bodily sensations. What may go wrong is the learned association: we may have associated a certain intrinsic sensation with a certain property (e.g., hotness), and then experience that sensation even when the property is absent.

But if, as Armstrong claims, bodily sensation *is* perception, what explains the difference in phenomenology between tactile perception and bodily perception? What explains the fact that sometimes the temperature may be attributed to the external object rather than the foot itself?

The answer Armstrong offers is that the tactile perception of heat is *dependent* on the bodily perception of heat. In bodily perception of heat, we experience a part of the body as having a certain temperature, and in tactile perception we perceive a *temperature difference* between the object and the body.

The discussion between Armstrong and Vesey ties into questions considered above: the status of thermoception as a somatosensory sense (or as bodily sensation, to use their terminology) precludes it from being a sense proper according to Vesey, who places it on par with itches and tingles. Armstrong on the other hand endorses thermoception as both an interoceptive and exteroceptive sense modality, where the interoceptive aspect is perception of temperature of body parts, and the exteroceptive aspect is perception of temperature differences between body parts and the environment.

In the next section I will discuss the distinction between tactile and haptic touch, which has bearing on the Armstrong-Vesey debate.

2.9 Tactile and haptic modes

Mohan Matthen has recently argued that touch has dual modes: a tactile mode that represents stimuli as happening at a location on the surface of the body, and a haptic mode that represents stimuli as being located in peripersonal space. (Matthen, 2021) An example of a purely tactile touch experience that Matthen provides is the ‘pins and needles’ sensation that can occur when you rest a limb in the same position too long. An example of a haptic touch experience is when you feel your phone vibrating in your hand.

Two characteristics distinguish tactile touch from haptic touch, according to Matthen:

1. Tactile touch is phenomenal and immanent, meaning that the properties experienced are attributed to the experience itself, and not to an external object. In the pins and needles example, the pricking sensation is attributed to the experience itself, and not to some object that impinges on the skin.
2. Tactile touch is located at a position on the body. If I move my leg while having a pins and needles sensation, the sensation moves with it.

Haptic touch on the other hand has the following corresponding characteristics:

1. Haptic touch is objective and transcendent, meaning that the properties experienced are attributed to external objects, not to the sensation itself. I feel the vibration as belonging to the phone, not to my hand. The experience is transcendent in the sense that it feels as though the vibration exists whether I feel it or not: the vibration is not dependent on my experience thereof. Haptic experience can be shared: someone else can feel the same vibration when touching your phone, which is not true for the pins and needles experience, according to Matthen.
2. Haptic touch is located in peripersonal space. Peripersonal space is the space immediately surrounding our bodies; the space in which we can grasp and manipulate objects. Peripersonal space can be centered on different body parts (head-centered, hand centered, trunk-centered). (di Pellegrino & Làdavas, 2015) When you move your hand relative to your vibrating phone, the sensation also

moves. This shows that properties are attributed not to an area of the body, but to a location in peripersonal space.

Haptic perception depends on tactile perception. We can attend to the tactile aspect of haptic perception, according to Matthen. In the case of the vibrating phone, you can attend to the specific sensation in our fingertips that occurs when you touch the phone. However, Matthen argues that haptic perception is not just a reformatting of tactile perception: haptic perception carries genuinely different information and is *emergent* from tactile perception (i.e., not reducible to tactile perception). Haptic perception emerges when the sensory system succeeds at integrating subsequent (or co-occurring) stimuli. The vibration of the phone felt in the fingertips is integrated with the shape, texture, weight, and hardness we feel the phone to have. Our sense of touch makes sense of the tactile vibrating sensation as belonging to a hard rectangular smooth object. In the case of the pins and needles sensation, the sensory system fails at making sense of the tactile sensation, so no haptic experience emerges. (Matthen, 2021)

In his discussion of the dual modes of touch, Matthen utilizes a broad view of touch that includes not only perceptions of vibrations or pressure or texture or shape, but also thermoception and pain. In fact, in most of the examples of tactile sensation that Matthen talks about, the sensation is a painful one. He says of pain and thermoception that they are sub-modalities that – on their own – do not give rise to haptic experience. It is only through integration with perceptions of vibrations and pressure etc., that pain experiences or temperature experiences can become (part of) haptic experiences.

I think it is no coincidence that Matthen's examples of tactile experience tend to include pain or thermoception. It is not only the case that touch allows for pain and temperature to be integrated into haptic sensations, as Matthen argues, it is also the case that painful and thermoceptive elements of a haptic experience allow us to attend to the tactile aspect of that experience. Berkeley, Ryle, Vesey, and Armstrong all agree that there is something peculiar about thermoception: that it is easy for us to distinguish the tactile element in experiences that include thermoception. Attending to the *purely tactile* aspect of a touch-experience is much easier if that experience includes pain or temperature sensation.

A possible explanation for the fact that pain makes it easy to attend to the tactile aspect of a sensation is that nociception has the two distinct elements of first pain and second pain that I mentioned in section 2.2.4. The second pain is felt 1-2 seconds after the first pain; rather a slow response.

Imagine someone drawing on the back of your hand with a ballpoint pen. If they press gently, you are likely to attribute the sensation of pressure to the pen. You are having a haptic experience. If you move your hand, the sensation of pressure will shift; you are locating the sensation not only tactually *on your hand* but also peripersonally as located in a hand-centric space.

Now imagine the person who is drawing on your hand pushes the pen down much harder, causing a painful sensation. If you withdraw your hand out of reach of the pen, the pain will linger on the back of your hand even when there is no contact. The second pain onset is delayed by 1-2 seconds, so the pain may still become more intense in the moments after withdrawing your hand. Your proprioceptive spatial awareness of the position of your hand relative to the pen is more easily integrated with the perception of innocuous pressure than it is integrated with the painful stimulus. I think this may be due to a timing mismatch: the decrease of pain simply comes too late to be connected to the withdrawal of the hand.

A similar argument could be made for thermoception: thermoreceptors are slower to respond than mechanoreceptors. Perhaps the delay between the motor command signal and the resulting perceptual stimulus makes it difficult to integrate these two signals. The same could be true for situations in which there is no motor command but where touch signals would be integrated with thermoceptive signals. The timing differences between the mechanoreceptor signals and the thermoreceptor signals could perhaps be relevant to how well these signals can be integrated into haptic experiences. This could possibly explain why it is easier to attend to the tactile aspect of haptic experiences that include pain or thermoception.

In the following chapter I will discuss the Heat Exchange Model of thermoception which, among other things, offers an alternative explanation for the dual modes of thermoception. I will present my own stance in the dual modes debate after discussing the Heat Exchange Model.

3 The heat exchange model of thermoception

One of the main questions of this thesis is what our experiences of hot and cold represent.[1] This chapter will begin answering that question. On a common-sense view, experiences of objects as hot or cold represent temperatures of (parts of) the environment. On another view, proposed by Richard Gray, they represent the exchange of thermal energy from the body to the environment or vice versa. This chapter evaluates three arguments that have been made for the second view. I argue that the common-sense view can accommodate these arguments just as well as the heat-exchange view. I conclude that the view that experiences of hot and cold represent heat exchange offers little explanatory value over the common-sense view.

In section 3.1 I introduce these two models of thermoception. Section 3.2 shows how both models account for the functioning of thermoreceptors. Section 3.3 is about the phenomenology of temperature experience, and whether it supports the view discussed. Section 3.4 revisits the dual modes of temperature perception discussed in the previous chapter and considers whether an explanation of the dual modes counts against either model of temperature perception.

3.1 Competing models of thermoception

What do we represent when we have experiences of objects as hot or cold? If we assume that in perceiving objects to be hot or cold, we form representations of the environment, then we can ask *which aspects* of the environment are represented by the sensory modality of thermoception.[2] What is the proper sensible of thermoception?

One straightforward answer is that we represent *temperatures* of (parts of) our environment. I will call this view the *Temperature Representation Model* (TRM) of thermoception.[3] On this view, an experience of hot or cold attributes a temperature to a certain object or part of the environment, or part of the body. It is not claimed by the TRM that the experiences represent temperatures as being on a certain scale: you don't necessarily experience the cup of tea as being 70° C or any specific numerical value for that matter. What the TRM claims is that thermoceptive experiences attribute a *temperature*, rather than

---

1 I would like to thank Richard Gray for very helpful comments on a draft of this chapter.

2 I will say a little more about anti-representationalist theories of temperature perception in chapter 4. For now, I am simply adopting a representationalist framework to compare two representationalist views.

3 In this chapter I am focusing on the thermal perception of external objects. An alternative version of the TRM could hold that temperature experiences represent the temperature of the skin, not of an object. In Chapter 6 I defend a representational theory of thermoception that claims we sometimes represent object temperatures, and sometimes skin temperatures.

some other property. Temperature would be defined in physical terms, as could be measured with a thermometer.[4] The TRM can be considered a common-sense view of thermoception.

The TRM has been criticized by several philosophers. Notably, Akins (1996) makes the case that thermoreceptors simply do not provide reliable information about temperatures in the environment or the temperature of (parts of) our bodies. Hence, it cannot be the function of the thermoceptive system to perceive temperatures. Akins uses this point to argue that thermoception is not a representational system at all, an idea I will discuss later in this thesis.

Another criticism of TRM is due to Gray (2013b). According to Gray there is an important mismatch between the character of temperature experience and the physical property of temperature. While temperature is a scale of increasing heat, thermoceptive experience appears discontinuous. Experiences of hotness are qualitatively different from experiences of coldness, and there is a middle zone where things feel neither hot nor cold. The physical magnitude of temperature does not have that same two-sided structure.

As an alternative to TRM, Gray adopts the *Heat Exchange Model* (HEM) of thermoception. According to this model, the thermoceptive system does not represent temperatures of the environment, but rather represents the exchange of heat with the environment. Experiences of coldness represent thermal energy being transferred from the body to the environment, while experiences of hotness represent thermal energy being transferred from the environment to the body.

The HEM as formulated in Gray (2013b) is about conscious experiences of external objects as hot or cold, although it can be extended to include unconscious perception and thermal experience of body parts. Indeed Gray (2023) presents an expanded view. Some of the arguments in this chapter deal with sub-personal representations – the presupposition is that those play an important part in bringing about personal-level thermal perception.

Both TRM and HEM hold that thermal experiences can yield perception, setting aside the error-theoretic option that thermal experiences massively misrepresent, and are not perceptual states. The notion of *experience* at play is one such that experiences can misrepresent: it is possible to have an experience of heat without there being actual heat.

The HEM explains the existence of a neutral zone or zero point of thermoceptive experience: in those cases, the system represents there being no or very little heat transfer in either direction. The theory also explains the qualitative difference between experiences of hotness and coldness: the difference of experience is due to the difference in direction of transfer. The transfer of thermal energy away from the body is experienced as cold and the transfer of thermal energy towards the body is experienced as warm. The difference in experience corresponds to a difference in the physical process (although this does not imply that temperature experience can be reduced to the physical process of heat exchange).

The HEM is also supposed to be in better standing given the functioning of the temperature receptors. It is a claim of the HEM that the receptors do not provide adequate

[4] At present it doesn't matter much which physical theory of temperature plays this role, but it could be fulfilled by e.g., a statistical mechanical description of temperature.

information for a veridical representation of temperature (c.f. Akins) but do provide the necessary information for a veridical representation of heat exchange.

Furthermore, the HEM is purported to provide a solution to the Armstrong-Vesey debate on the dual modes of thermoceptive experience (discussed in the previous chapter). Both Armstrong and Vesey hold that the two modes are different in terms of what they represent. (Armstrong, 1962, 1963; Vesey, 1960) The HEM provides a way of saying both modes of thermoceptive experience represent the same quantity. That is: the exchange of thermal energy to or from the body.

In short, the heat exchange model purports to be superior to the competing temperature representation model on the basis of three arguments:

1. The way temperature receptors work makes the TRM implausible but provides support for the HEM.
2. Temperature is one-dimensional, while thermoceptive experience is two-dimensional and has a zero point (that isn't the absolute zero of the temperature scale). The HEM explains this, while it would be an odd phenomenon for the TRM.
3. Feeling something to be hot with a certain body part is different from that body part feeling hot. The HEM allows us to see both experiences as having the same type of content.

In this chapter I argue that these three arguments do not provide adequate reason to think the HEM is true. Against the first argument I will bring evidence that while the thermoreceptor response might indeed not allow for accurate representations of temperatures, it doesn't allow for reliable representation of heat exchange either. The very same arguments that count against the TRM also count against HEM. Against the second point I argue that a two-dimensional structure with a zero point around body temperature need not mean the system isn't representing temperature. In fact, it makes sense for a system that is aimed at maintaining homeostasis to be structured this way. The question of *what is represented* isn't answered by appealing to the two-dimensional nature of thermoceptive experience. Against the third point I argue that the proposed solution is not only available to a proponent of the HEM, but also to a proponent of the TRM.

3.2 Thermoreceptors and the two models

The first argument under scrutiny is that the HEM fits better with the way thermoreceptors function. Recall from chapter 2 that there are four main functional kinds of receptors in the skin associated with thermoception. There are two varieties of nociceptors, for noxious heat stimuli and noxious cold stimuli respectively. The firing of these receptors for extreme heat or cold results in experiences of pain. Then there are two kinds of receptor for hot and cold respectively for the innocuous range of temperature.

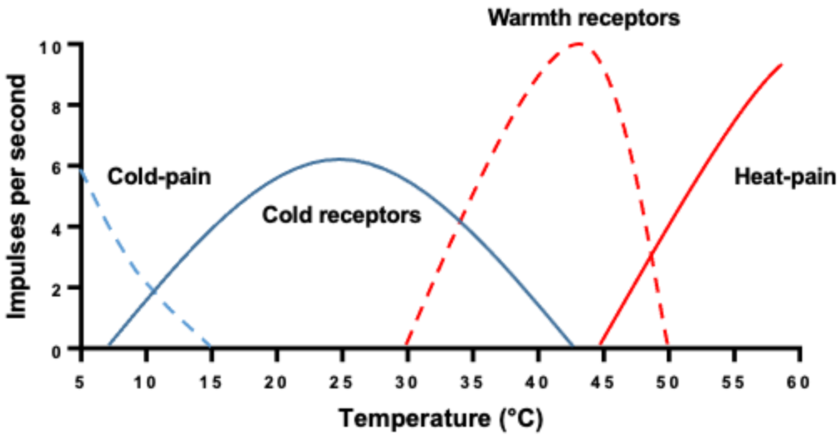


Figure 4: Receptor responses for static temperatures. Adapted from Hall & Guyton (2011).[5]

The receptors for innocuous cold and innocuous warmth have bell-shaped response curves (see fig.1). The static response curves of these innocuous hot and cold receptors partially overlap, and the crossover point lies near our normal body temperature.

Because of these receptor responses, Akins argues that there can be no reliable correlation between the response of the receptor and a stimulus temperature. For example, a skin temperature stimulus of 21°C and 29°C would elicit the same response from the cold receptors. At both these temperatures the firing rate of the specialized innocuous cold fibers would be equal. The bell-like shape of the response curve of individual receptor types results in equivocation of temperatures: for a reliable correlation between stimulus temperature and neural response, there would have to be a "computable function" relating the two, and there simply is not. (Akins, 1996, p. 357)

The argument Akins makes also counts against HEM. Holding equal the body temperature of the individual, the area of skin that is stimulated, and the thermal conductivity of the probe, each stimulus temperature corresponds with a certain amount of heat transfer to or from the body. Lower temperature stimuli result in more heat being transferred away from the body. If cold receptors are meant to reliably indicate the amount of heat transferred away from the body, we would expect to see an increased response to lower stimulus temperatures. But as you can see, cold response only increases until about 25°C, and then gradually drops off. There is a cold receptor response rate that equivocates stimulus temperatures of 21°C and 29°C, but similarly there are two rates of heat exchange that elicit the same cold receptor response.

Akins considers several possible responses to this criticism, including the idea that the combination of the various receptors may provide reliable correlation with temperature. Akins brings against this that there is great variation across the body of relative concentrations of warm vs. cold receptors. The combination of signals would have to take

[5] Receptor response research: (Darian-Smith et al., 1973; Dubner et al., 1975; D R Kenshalo & Duclaux, 1977)

into account the relative concentrations of receptors at specific locations, and there is no reason to think that information is available to the central nervous system.

Furthermore, there is a difference in the absolute concentration of thermoreceptors across the body, irrespective of type. Some areas of the body are more sensitive to thermal stimuli than others. The same stimulus applied to different parts of the body elicits a different response in the neural population. (Nadel, Mitchell, & Stolwijk, 1973; Stevens & Choo, 1998) For example, the hand is more densely innervated with temperature receptors than the thigh or shin, so that a cool stimulus on the hand would elicit a stronger cold response in the neural population than the same stimulus when applied to the shin. This, again, counts as an argument against both TRM and HEM, since both are representational theories that depend on there being a regularity between stimuli and receptor responses.

To make matters worse for both TRM and HEM, thermoreceptors exhibit an exaggerated response to temperature *change.* When the skin is actively cooled, cold receptors will exhibit a strongly increased firing rate, followed by adaptation to the normal static response. When the skin is warmed, activity is temporarily inhibited, followed by adaptation to static response. Warm receptors conversely show an exaggerated response to warming, and inhibition with cooling. (Schepers & Ringkamp, 2010)

The dynamic response to temperature changes once more shows that receptor response is not reliably correlated with temperatures. One temperature may elicit different responses when presented as part of a dynamic stimulus. Additionally, a single firing rate may be the consequence of different temperatures, depending on whether they were presented statically or dynamically.

This poses a problem not only for TRM, but also for HEM. From the dynamic response curves, you can see there are points with equal probe temperature that give different receptor responses. The proponent of the TRM must conclude that the receptors exaggerate temperatures during change. But those points in time also have equal rates of heat exchange, so the proponent of the HEM must conclude that the receptors exaggerate change in the amount of heat transferred.

A proponent of the heat exchange model could follow a different line of argument: the dynamic response to temperature change can be explained as the predictable behavior of receptors attuned to heat exchange. A faster rate of heat exchange away from the body is associated with the skin losing heat quickly. So, if the skin is actively *cooled,* we expect the system to represent massive heat transfer away from the body. However, this leaves open the question why this dynamic response is transitive. As the figure above shows, the dynamic response abates after a short period of time (a matter of seconds), and firing rates return to normal static levels. If the system is functioning properly when it represents cooling as massive heat transfer away from the body, why would it adapt?

The proponent of the Heat Exchange Model may be inclined to say that cool adaptation is a consequence of heat being transferred from the body to the object, effectively warming the object, or colling the skin to a point of thermal equilibrium. However, the probe with which such stimuli are applied in thermoception experiments is normally temperature-controlled by running water of a constant temperature through it. This prevents the body temperature of the subject from significantly warming or cooling the stimulus probe. The adaptation occurs on the neurological side and is not due to a physical change in temperature of the probe. This can also be seen from the fact that different nerve fiber types involved in thermoception have adaptation characteristics: some fibers adapt more quickly than others. This is not a consequence of those fibers reaching thermal equilibrium before other fibers do, it is a neurological phenomenon (see also section 2.2).

Another aspect of receptor response that is cited as a support for HEM is that there is summing of the response, depending on the size of the area that is stimulated. (Gray, 2013b) When a larger area of the body receives a warm stimulus, there is a stronger receptor response, and similarly for cold stimuli. (Johnson et al., 1973; Dan R Kenshalo, Decker, & Hamilton, 1967). In many cases the increase in size is not perceived as such, but only as an increase in temperature. (Stevens, 1991) If the goal of the thermoceptive system is to represent the temperatures of objects, why would a larger object feel hotter than a smaller object with the same temperature? And why would there be such poor spatial perception? These questions pose a problem for the proponent of TRM. For the HEM, these results are less problematic. The total heat transferred to or from an object depends on that object's temperature, heat conductivity, the contact area, and the exposure time. In the spatial summing case where the object is not perceived as being larger but only as being colder or hotter, the size is misrepresented, but the rate of heat exchange could still be accurately represented.

A similar point holds for varying the thermal conductivity of the stimulus. We experience thermally conductive stimuli that are colder than skin temperature as being colder than less conductive stimuli of the same temperature. The HEM easily explains this: the conductive object results in quicker heat transfer. The TRM has more difficulty accounting for it: it must be seen as a case of misrepresentation since both stimuli (conductive and non-conductive) are the same temperature. In chapter 6 I will attempt to explain both spatial summing and the conductance effect as a *useful* type of misrepresentation when we construct a more sophisticated version of the TRM. The points on spatial summation and conductance do count in favor of HEM over the simple TRM.

The underlying reason why the basic tenet of Akins's criticism against the TRM also counts against the HEM is simple. The exchange of thermal energy between the body and the environment is usually determined by the temperature difference between them. In experimental setups where temperature is the variable, and other factors such as material and the size of the stimulus probe are constant, the exchange of heat is nomically covariant with the stimulus temperature. Under those circumstances, what counts as an argument against TRM automatically counts as an argument against HEM. In experimental setups where the stimulus size is a variable (i.e., spatial summing experiments) the predictions of the HEM and TRM come apart. In explaining the results of spatial summing cases, HEM has the edge over TRM. But the lack of reliable covariation remains a problem to be explained by both theories.

Polymodality of receptors

As described in Chapter 2, some thermoreceptors can activate in response to stimuli outside what we think of as their main response function. Cold fibers not only react to their temperature or heat exchange stimuli, but also can react to other chemical stimuli, menthol in particular. (Bautista et al., 2007; Chuang, Neuhausser, & Julius, 2004; Hensel & Zotterman, 1951a) Menthol doesn't inherently have a low temperature, and it doesn't cause any special heat exchange. Both TRM and HEM will have to grant that the cold sensation caused by menthol is a case of misrepresentation.

Secondly, some cold receptors are also activated by temperatures in the noxious heat range. (Campero, Serra, Bostock, & Ochoa, 2001; Dubner et al., 1975; D R Kenshalo & Duclaux, 1977; Long, 1977) When cold fibers are stimulated with noxiously high temperatures, a 'paradoxical' cool sensation can occur. The paradoxical sensation should count as misrepresentation on both TRM and HEM.

Thirdly, there are many receptors that respond to temperature or temperature change while their primary function seems to be otherwise. Cold receptors themselves are not sensitive to mechanical stimulation, but about half of all slow-adapting mechanoreceptors respond to cooling. This mechanism likely underlies the phenomenon known as the Silver Thaler Illusion or the Weber deception. The illusion is that cold stimuli can evoke a sense of pressure: a silver coin will feel heavier when it is cold. (Cahusac & Noyce, 2007; Duclaux & Kenshalo, 1972; Hensel & Zotterman, 1951b; Iggo & Muir, 1969; Tapper, 1965)

In this case, both the TRM and the HEM are misrepresenting pressure – there is no significant misrepresentation of temperature. The fact that thermal stimuli affect pressure sensation is then seen as a quirk of the mechanoreceptive system, not as a feature of the thermoceptive system. This is supported by the fact that the response of these mechanoreceptors to cooling is much smaller than to mechanical stimulation, and also much smaller than cold fiber response. (Johnson et al., 1973)

Warm fibers are also to some extent polymodal. The molecular mechanisms for warmth reception (TRPV3 and TRPV4)[6] are sensitive to some chemical stimuli. In fact, TRPV3 may be involved in flavor perception. (Xu, Delling, Jun, & Clapham, 2006) TRPV3 may also play a role in noxious heat perception. Camphor, applied to human skin, causes a sensation of warmth. (Moqrich et al., 2005) TRM and HEM must count this as a thermoceptive illusion.

To conclude, the responses of thermoreceptors show us that there is no straightforward correspondence between their neural activity, and temperature or heat exchange. The variable distribution of thermoreceptors further exacerbates those worries. Spatial summing effects provide some measure of support for HEM over TRM. The argument that the HEM matches better with the way thermoreceptors function has limited validity. If anything, the way thermoreceptors function suggests that thermoception does not represent any single physical quantity at all.

[6] The role of TRPV3 and TRPV4 in temperature perception is contested. See for example (Huang, Li, Yu, Wang, & Caterina, 2011)

3.3 Phenomenological structure of thermoception

The second argument I am concerned with is the mismatch of the phenomenology of thermoception with the structure of the domain of temperatures. The argument is that temperature, as a physical quantity, is a magnitude with a single dimension.[7] Thermoceptive experience, on the other hand, has two distinct ways of presenting: as *warm* or *cold* If thermoceptive experiences represented temperatures, why would two ranges of the scale have distinct qualitative character? If we accept that thermoceptive experiences represent heat exchange, the qualitative difference makes more sense: at temperatures below our own body temperature, there is negative heat exchange (body-to-environment). At temperatures above our own body temperature, there is positive heat exchange (environment-to-body). On the HEM there is a neat match-up of three elements: the physical quantity of heat exchange, the response curves of the warm and cold fibers, and the phenomenology of thermoceptive experience.

How strong is this duality of phenomenology? A proponent of the HEM may claim that phenomenologically, hot and cold are two distinct dimensions: a cold stimulus does not feel *less hot*, it just feels entirely different from a hot stimulus. I have some doubts about that phenomenological claim. Warm experiences can be ordered from less warm to warmer, and cold experiences can be ordered from less cold to colder. But I think it is also the case that experiences of cold are perceived to be *less warm* than experiences of warmth. Warm and cold experiences are not two independent axes, they are structurally related by reference to the objective temperature. Does phenomenology really suggest we aren't tracking a single continuous quantity? I do not deny that there is a significant qualitative difference between the experience hot and cold, but I am not convinced the 'structure' of experienced temperature doesn't put these distinct experiences on the same gradient scale. Ultimately this is a clash of intuitions, and the only way I see of settling this disagreement is a systematic investigation of temperature phenomenology.

Furthermore, the proponent of TRM may respond that the dual phenomenological structure of temperature experience is not surprising, given an extra assumption about what thermal experience *does* for the organism. It is a plausible assumption that the main function of conscious experiences of innocuous temperatures is to play a part in behavioral thermoregulation.

The main function of the thermoceptive system is to provide inputs for thermoregulation. A big part of this is done autonomously: when your core temperature is low, you shiver. When your core temperature is high, you sweat. But some thermal situations are to be handled with behavior: dependent on thermal conditions animals will adapt their stance, position towards the sun, find shade, huddle together, or even migrate to warmer climes. Most animals have a range of body temperatures in which they can function that is much narrower than the range of environmental temperatures they are exposed to. Behavioral responses to temperature are important for maintaining homeostasis.

---

7 For an in-depth look at the mathematical structure of the physical quantity of temperature, see Skow (2011).

On this picture of the *goal* of the thermoceptive system, it makes sense even according to TRM that the receptors and phenomenological structure of thermoception are the way they are. If homeostasis is the end goal of the system that consumes temperature representations, it makes sense that there is a neutral zone in which there is little thermoceptive activity, since this stimulus does not require a response. It also makes sense that the phenomenology of hot and cold is different: the appropriate behavioral response to heat is different than the appropriate behavioral response to cold.

If the function of the system is to represent *heat exchange*, we can expect this match-up of two kinds of receptors, and two kinds of experience. But if the function of the system is to represent temperature with the aim of maintaining thermal homeostasis, then we can also expect this match-up of two kinds of receptors and two qualitatively different experiences.

Another, much simpler argument can be made against the phenomenological claim of HEM. For most humans, our normal environmental temperature lies well below the point at which there is no heat exchange between the body and the environment. Zero heat exchange is achieved when the environment is the same temperature as the body (i.e., around 37°C). To most people, an environment of 37°C feels decidedly hot, probably uncomfortably so. From the perspective of thermoregulation, this makes sense. As endotherms, humans produce most of their body heat themselves; a significant amount of heat is produced by normal muscular activity. In an environment of 37°C, normally the thermoceptive system will be increasing thermolysis through sweating and vasodilation, to avoid hyperthermia. If there is such a thing as a 'neutral' temperature for humans, it is much lower than body temperature. Most of the time, humans need to get rid of a lot of self-produced heat.

The fact that the crossover point for the static response curves of hot and cold thermoreceptors lies around body temperature may be responsible for the curious phenomenology of touching something that is around 37°C,[8] but that doesn't mean that this point represents 'neither hot nor cold'. When it comes to perceiving environmental temperatures, this 'neither hot nor cold' point lies much lower. The HEM is primarily a theory of the perception of temperature of external objects (and radiant heat), and perhaps not primarily concerned with environmental temperatures. But is such a strong theoretical division between perceiving object temperatures and environmental temperatures a good idea? If the HEM were to apply to environmental temperatures, and heat exchange away from the body was perceived as cold, then we would (almost) always be feeling cold.

The proponent of the HEM could reply that the interoceptive system that regulates core temperature is not a part of the sense modality of thermoception. However, that argument is somewhat undermined by HEM itself: while TRM claims that thermoception can represent the temperatures of external objects, HEM only allows for representations of transfer of heat. This makes HEM skew decidedly more interoceptive (although it is not purely so), and to claim that thermoception does not include the interoceptive thermal information gathering

---

[8] A phenomenology I personally have trouble replicating. The 'neither hot nor cold' sensation supposedly felt at that temperature doesn't stand out to me as qualitatively different from touching other things that are above room temperature.

that supports thermoregulation seems somewhat arbitrary. Especially when one considers that shell temperature information feeds into central thermoregulation. The same receptors that allow you to feel an object as hot or cold, also contribute to central thermoregulation. Even on a version of the HEM that does not include interoceptive thermoregulatory signals, it would seem like the neutral zone of temperature experience just does not match up to the neutral zone of heat exchange. What we feel as a neutral or comfortable environmental temperature is in fact a transfer of heat away from the body.

3.4 The dual modes of thermoception

The third argument for the HEM is that it settles the Armstrong/Vesey debate about the two distinct *modes* of thermoceptive experience. If an object feels warm to the touch of the hand, then the warmth is felt in the hand. But there is a difference between your hand feeling warm and an object feeling warm to your hand. And in some cases (for example: fever) a part of our body may feel cold in and of itself, while it is warm to the touch of a hand.

Vesey's explanation of the related contents of these two modes of sensation is that it is the association with a previous perceiving sensation of hotness, that makes the intrinsic sensation one of hotness. The peculiar duality of temperature perception then is explained by the activity of two distinct processes, one perceptual and one not. The non-perceptual process of *intrinsic sensation* acquires its 'hotness' content through learned association with the perceptual process of *perceiving sensation*. (Vesey, 1960)

Armstrong's competing explanation of the two modes of temperature experience is that what he calls *bodily perception* and *tactile perception*[9] are distinct, but both are forms of heat perception. The tactile perception of heat is dependent on the bodily perception of heat. In bodily perception of heat, we experience a part of the body as having a certain temperature, and in tactile perception we perceive a *temperature difference* between the object and the body. On both Armstrong's and Vesey's accounts the two modes of temperature experience represent different things. (Armstrong, 1962, 1963)

The heat exchange model also offers a perspective on this discussion: according to the heat exchange model, both the perception of one's own body as warm or cold, and the perception of an object as warm or cold are cases of representing heat exchange. What makes the perception of the temperature of an external object different from perception of the own body is the addition of touch. The idea, originating with Strang (1960), is that touch experience is a necessary condition for the experience of *an object* as warm or cold. When we feel the warmth of the sun on our skin, we do not attribute this warmth to the object (the sun) but rather to our skin itself, precisely because there is no touch experience that accompanies the warmth.

[9] What Armstrong calls bodily perception aligns more closely with what Matthen calls tactile experience and what Matthen calls haptic experience is closer to what Armstrong calls tactile experience.

According to the HEM, the added tactile information given by physically touching something does not change the content of the representation of warmness: the content is still the exchange of heat between the body and the environment.

This strategy, however, is not uniquely available to HEM. A proponent of TRM can agree that tactile experience is a necessary condition for the experience of *an object* as warm or cold. Thermoreceptors are not sensitive to mechanical stimulation, so the thermoceptive system itself does not have the tools to distinguish between a certain part of the skin being cold, and something that's touching the skin being cold. The proponent of TRM should say that the experience of external objects as warm or cold is a product of cross modal spatial attribution. The temperature sensation in a certain part of the skin is attributed to an object *because* of tactile stimulation in the same area.

Note that this proposal is essentially the same as Matthen's theory of the dual modes of touch. The experience of thermoception *without* touch is what Matthen would call a *tactile* experience, whereas thermoception *with* touch has the possibility of becoming a *haptic* experience. If we translate Strang's theory into more modern terminology, his point is that binding to touch is necessary for thermoception to attribute properties to objects in peripersonal space. Without binding to touch, there is no object-temperature perception. Matthen seems to make the same point in his remarks about the role of thermoception in the dual modes of touch: thermoception on its own does not allow for haptic experience. Whether Matthen thinks this ascribes a *temperature* to the external object, or some other thermal property, is not entirely clear. Yet a proponent of the HEM would say that the haptic mode of thermoception ascribes an exchange of heat with an object, not a temperature of the object. The option of a Matthen/Strang style theory is certainly also open for a proponent of TRM, and I will present such a theory in chapter 6.

A parallel may be drawn with flavor perception. Flavor experience involves more than just chemoreceptors on the tongue. Retronasal olfaction is of great importance to flavor perception. These smell receptors, which contribute a large part of the flavor experience, are located in the nasal cavity, in the epithelium. Yet, when eating, flavors are experienced as being in the mouth. The spatial attribution of retronasal olfaction to the mouth (known as *oral referral)* is mediated by flavor intensity on the tongue. To put it simply, in flavor chemoreception on the epithelium is *bound together* with chemoreception on the tongue and spatially co-located by this binding. (Auvray & Spence, 2008; Spence, 2016)

Object-temperature perception similarly depends on integration of information from different sources. Touch mediates the binding of thermoceptive stimuli to objects in peripersonal space. This option is open to both the HEM *and* the TRM, so the dual modes of thermoception do not provide an argument for one model over the other.

3.5 HEM and the sense of touch

The Heat Exchange Model treats thermoception as separate from touch, and it also constitutes an argument for doing so. In chapter 1 I have discussed criteria of sensory individuation, and whether thermoception is part of the sense of touch. I have omitted one argument for the distinctness of touch and thermoception in that chapter, because it depends on the HEM. Gray (2023), following Fulkerson, categorizes sense modalities as *distal* senses, that allow us to perceive objects at a distance or *proximal senses*, that only allow us to perceive objects that are touching us. Touch, according to Fulkerson and Gray, is neither: one can perceive distal objects through touch, but only if there is an appropriate tactual medium present. For example, we can feel the ground under our feet through our shoes, we can feel the shape of an object through gloves, and we can feel the bumpiness of the road through a car. This makes touch a *connection sense*.

Thermoception, according to Gray, is not a connection sense. This is easily argued if one takes the HEM to be true: temperature experiences simply don't represent the thermal properties of external objects, they only represent the heat exchanged to or from the body. When you touch a hot or cold object *via* a thermal medium, you would only represent the heat exchange with that medium, and not with the distal object.

Gray also argued that in the case of perceiving an object that emits radiant heat (in his example, a candle), we can perceive the heat to come from a certain location, but we are not engaged in distal perception proper, because according to the HEM we are not perceiving any property of the distal object, but only the transfer of heat to or from the body and some spatial content as to which direction this heat is coming from.

The final step of the argument is that thermoception cannot be part of touch *because* it isn't a connection sense: the categorization of senses as distal, proximal, or connection senses apparently does some work in individuating the senses.

I disagree with much of the argument, but not with its conclusion. I have argued that thermoception can be regarded as separate from touch based on the modified neuroethological account. I think the argument Gray (2022) makes depends for much of its strength on the assumption that the HEM is true. If one is inclined to think that thermal touch attributes thermal properties to external objects, then the difference between touch and thermoception seems less stark.

Another problem I have with the argument is that it depends on the notion of senses being essentially distal or proximal or connection senses, and that this is an individuating principle. The difference between proximal senses and connection senses is one of psychological attribution: a connection sense reacts to stimuli at the edge of the body, but in some cases projects that information onto distal objects. The relation between sensory individuation and sensory attribution is not self-explanatory, and the argument from HEM to the distinctness of thermoception essentially depends on this relation.

3.6 Conclusion

It is alleged that the view of thermoception as reliably representing temperatures is made implausible by the details of the functioning of thermoreceptors. The view of thermoception as representing heat exchange to or from the body faces considerable difficulties for the same reason. If representation requires there to be a reliable correlation between receptor response and the quantity represented, both views possibly fail.

The idea that the phenomenological structure of temperature perception matches better with the heat exchange view is doubtful. On a plausible picture of the function of thermoception, the phenomenology makes sense even on the HRM.

The heat exchange model, combined with an appeal to tactile perception, provides an answer to the debate on the dual modes of thermoception. However, the appeal to tactile perception has the same implication of deciding the debate when one believes thermoception represents temperatures rather than heat exchange.

I hope to have shown these three arguments for HEM over TRM can be countered. In the following chapters I will be arguing for a TRM-like theory of thermoception, albeit a more sophisticated one than the simple view used in this chapter.

4 Representations

In this chapter I outline a theory of representations that can help characterize thermoception as a representational system and provides a way of thinking about the content for such a system.

In section 4.1 I outline the main questions concerning mental representation. In section 4.2 I briefly touch on theories that deny that perception is representational. Section 4.3 deals with accounts of mental content that, unlike mine, do not aim to *naturalize* content. Section 4.4 introduces the framework of an informational theory of representation, and section 4.5 explains how biological functions play a constitutive role in representational systems. Section 4.6 explicates the notion of a task function, which is essential to my preferred theory of representation. Section 4.7 deals with a debate between producer-based and consumer-based accounts of representational content. The notion of task functions that I use puts me on the side of consumer-based accounts. Section 4.8 puts together the notions of a task function with the framework of an informational correlational theory of representation, which gives us a fully-fledged theory of mental representation. Section 4.9 explores an exciting addition to 'standard' informational accounts of representation, that holds that representations are states of informational systems that realize a sweet spot in representing faithfully with minimal resources.

4.1 Representational Status and Content

It is useful to distinguish between two questions about mental representation. The first question asks what makes it the case that a particular mental state functions as a representation. This is known as the *status question* because an answer to it determines which mental states have the status of being representations. The second question is what makes a particular mental state (or system) a representation of *A* rather than *B.* This is known as the content question because an answer to it determines whether the content of the representation is A or rather B. (Ramsey, 2016)

General theories of representation give general answers to these questions: a set of conditions that differentiates all representations from all non-representations and a set that explains why individual representations represent what they represent. Such general answers have consequences for specific classes of mental states. Do linguistic thoughts count as representations? Do conscious perceptions count as representations? Does a certain pattern of neural activity in the primary visual cortex (V1) count as a representation? And what exactly does that pattern of activity in V1 represent?

The disagreement between the HEM and TRM discussed in the previous chapter amounts to different answers to the content question. Both theories agree that thermoceptive experiences count as representations, so regarding this class of mental states, they agree on the status question. But they disagree on what exactly is represented: TRM claims thermoceptive states represent a temperature, while HEM claims they represent a rate of thermal energy exchange. This shows that giving answers to these very general questions about representation influences what one should think about specific questions concerning perceptual content.

The general theory of representation that I adopt is an informational theory of mental representations. The main idea is that mental states can function as representations of something because they carry information about this thing, if it is part of a system in which the information is relevant to the successful functioning of the system.

4.2 Wholesale denials of representational status

Before going into the theory of representation, I want to shortly mention the view that perceptual states are not representations. Representationalism about perception is the idea that perceptual states are representational states, and that perception works the way it does *because* those states are representational states. It amounts to giving (or assuming) an answer to the status question as it pertains to perceptual states. Although representationalism is the dominant position in philosophy and in perception science, several anti-representationalist theories have significant support.

The behaviorist paradigm in psychology of the 20th century is broadly anti-representational. Behaviorism can be understood as a normative claim about the subject matter of scientific psychology. On this understanding (*methodological behaviorism*) psychology should not be concerned with inner mental states, but only with behavior. (e.g., Sellars, 1963; or the work of John Watson) Representations are arguably such inner mental states (they are certainly not behaviors), so in behaviorist psychology there is no room for mental representation.

If behaviorism is understood as the claim that mental states do not figure in the ultimate explanations of behavior (*psychological behaviorism*), then it denies the explanatory relevance of mental representations.[1]

If behaviorism is understood as the claim that mental terms in psychological explanations need to be translated into or analyzed in behavioral terms (*analytical behaviorism*), then there should be some behavioral analysis of the concept of mental representation and talk of representations would be superfluous or simply shorthand for behavioral explanations.[2] All three types of behaviorism are incompatible with a psychology which uses mental representations as genuine units of explanation.

One of the philosophical draws of behaviorism was that it was a naturalist theory. The approach to mental representation that I take in this thesis is meant to be compatible with naturalism, allowing for non-intentional explanations of intentional content. Behaviorism in

---

1 Psychological behaviorism was a dominant research paradigm in the 20th century, in the work of such psychologists as Pavlov and Skinner.

2 See for example Ryle (1949) and U.T. Place (2004)

all three forms has been surpassed by a psychology that does allow for genuine mental states on empirical grounds.

Approaches that emphasize the enactive aspects of mental processes sometimes include a rejection of representationalism. These approaches hold that cognition is importantly dependent on continuous interaction between the body and the environment. This is thought to be in contradiction with a 'classical' representationalist view of cognition as computation over symbolic representations. The supposed incompatibility of this classical representationalist framework with those enactive aspects of perception leads some philosophers to deny that perceptual states are representations. (Hutto & Myin, 2013; O'Regan & Noë, 2001)

Another group of anti-representationalists are those who think perceptual states are not representations separate from the things they represent, but relations between the observer and the object being observed. On these views (referred to as *disjunctivism* or *naïve realism)* a perceptual state doesn't represent the object, rather the actual object is a constitutive element of the perceptual state. On this view misrepresentation is impossible, because there is no perception if the object does not figure in the relation. On a disjunctivist view, only veridical perceptions are perceptual states. False perceptions are states of some other kind.

I will not engage with these anti-representationalist theories. The first reason not to do so is that it would simply be too far removed from the main topic of this thesis. The second reason is that my arguments for representationalism are unlikely to convince dyed-in-the-wool enactivists or disjunctivists.

What I will offer instead is a type of representationalism that I think will be less objectionable to some anti-representationalists. Current theories of representation do not necessarily have the same commitments as 'classical' theories of cognition. Specifically, a rather minimal notion of representation doesn't come with the same demands on cognitive architecture that 'classical' representationalism does.

4.3 Non-naturalist approaches

An important motivation behind accounts such as informational theories of content and teleosemantics is to explain mental content, which exhibits intentionality, in non-intentional terms. The idea is that intentionality is a non-natural feature of the world, and to have a theory of mental content that conforms to naturalism, one needs to spell it out in non-intentional terms. Of course, it is possible to decline to do so. One option is to simply leave the intentionality of mental content unexplained. The goal of this thesis, however, is to explain the content of certain types of mental states, and how they came to have this content.

Another option is to give an account of mental content, but not to try to explain it in non-intentional terms. The most prominent of such accounts is a theory known as *phenomenal intentionality.* On this theory, the intentionality of mental states is explained in terms of phenomenal consciousness. A central claim is that phenomenal consciousness is explanatorily or metaphysically prior to mental content. According to phenomenal intentionality, the intentional content of a state is explained by its phenomenal content. Take a perception of a sound as having a certain pitch, middle C for example. The idea is that the content of this perception derives from the phenomenal experience of a middle C. There is a phenomenal state that is *what it is like* to experience a middle C, and this perceptual state has that character, and therefore it represents the sound as having that pitch. Some versions of phenomenal intentionality claim *all* intentional states *are* phenomenal states. (Farkas, 2008; Pitt, 2004) Weaker versions claim that intentional states are *grounded* in phenomenal states. (Loar, 2017b, 2017a; McGinn, 1988; Searle, 1983) Some theorists also restrict the scope of the claim: some but not all intentional states derive their content from phenomenal states.

I have two main reasons not to support phenomenal intentionality. The first reason is that it doesn't provide an answer to how mental content is compatible with a naturalistic worldview. At best, phenomenal intentionality merely relocates the problem of naturalizing mental states from intentional states to phenomenal states: explaining phenomenal consciousness in naturalistic terms will be at least as hard as explaining mental content. So, the problem simply moves to another domain. At worst, phenomenal intentionality is actively anti-naturalistic and claims that intentionality *cannot* be explained in non-intentional terms.

The second reason not to support phenomenal intentionality is empirical: while perception *can* be conscious, it isn't always so. In humans, a lot of perceptual information is processed unconsciously, even though we have a capacity for conscious perception.

A remarkable example of unconscious perception is *blindsight*, a neurological condition where a patient does not possess phenomenal experience of visual perception (often in one half of the visual field), while retaining the ability to act on visual stimuli. (Weiskrantz, 1990) But unconscious perception isn't just present in subjects with neurological conditions, it is common in healthy subjects too. Visual processing happens in two main pathways: the ventral stream and the dorsal stream. Dorsal stream processing is essential for the visual guidance of many motor actions (e.g., grasping), yet is largely unconscious. This shows that perceptual input necessary for sensorimotor control need not be conscious. (Goodale & Milner, 1992, 2013; Kravitz, Saleem, Baker, & Mishkin, 2011; Ungerleider & Mishkin, 1982)

If the perceptual states that guide motor action are intentional states, then the strongest form of phenomenal consciousness, which claims *all* intentional states are phenomenal states, must be false.

Weaker versions of phenomenal intentionality, that acknowledge that not *all* intentional states are phenomenal states and claim intentionality is grounded in phenomenal consciousness may be able to explain cases of unconscious perception in humans – while those particular intentional states are not identical to phenomenal states, their contents depend on phenomenal consciousness in some other way. But a view of this type overlooks the possibility of organisms that have perceptual abilities, but no phenomenal consciousness. In my view of perception as outlined in the previous chapters, there is a continuity between the simplest responsiveness to aspects of the environment and the most complex conscious perceptual processes. Perception is widespread in the animal kingdom and possibly something perception-like is present in other branches of the tree of life. It is plausible that phenomenal consciousness requires a greater deal of neurological complexity in an organism than perception does. If this is right, how are the contents of the perceptual states of these 'simple minds' grounded? Plausibly, perception is evolutionarily prior to consciousness: why would consciousness exist in an organism with no perceptual input?

4.4 Informational theories of representation

On an informational theory of representation, a state represents something because it carries information about it. An informational theory of representation answers both the status question and the content question: a state can function as a representation of A because it carries information about A.[3]

What is it for a state to carry information? The idea is that a state (or property or object or process) carries information if it correlates with some other state (or property or object or process).[4] Think about a car alarm: it is meant to go off when the car is being broken into. A perfect alarm always goes off when the car is being broken into and never goes off when the car isn't being broken into. This perfect alarm can be used to obtain perfectly reliable information about whether the car is being broken into. Now think of a car alarm that malfunctions a bit. Sometimes it may go off when the car *isn't* being broken into, and sometimes it may malfunction by *not* going off when the car *is* being broken into. If the car alarm goes off completely randomly, with no correlation between the state it is supposed to represent and the resulting alarm, then it is useless for telling whether your car is being broken into. If there is malfunction *sometimes*, but there still is correlation, then the car alarm can still be useful, as long as the alarm going off raises the probability that your car is being broken into, or the alarm *not* going off raises the probability your car is not being broken into. Biological representational systems are usually a bit like the slightly malfunctioning car alarm: they can signal a state, but they tend not to do so without error.[5]

If we were to formalize the representational state as some object or process a being in the state $F,$ and the represented state as some object or process b being in the state G, then we could define carrying correlational information as follows:[6]

> Correlational information:
> Fa carries correlational information about Gb iff $P(Gb|Fa) \neq P(Gb)$

If the above conditional holds, then Fa can be used to gain information about Gb. But there is a problem with this definition as it pertains to biological systems carrying information.

---

[3] Informational theories of representation in the philosophy of mind were pioneered by Dretske (e.g. 1981).

[4] I will simply talk about states from now on, but that should not be taken to exclude the option that representations are dynamic processes. In fact, I think it depends on the specifics of the representational system whether it is better to talk about states, properties or processes doing the representing, and about whether it is states or properties or processes that are being represented.

[5] The basic idea that underlies all this is that a message that contains information reduces uncertainty. The measure of the uncertainty reduced is the measure of the amount of information in a message. This quantity is known as (informational) entropy. (Shannon, 1948)

[6] All definitions in this section are quoted or slightly paraphrased from Shea (2018)

There is the possibility that a correlation is merely accidental and doesn't track a probability that is nomologically connected. For example, my car alarm could be malfunctioning; it spontaneously goes off every Friday night, whether the car is being broken into or not. Now it so happens that there is a car thief active in my neighborhood every Friday night and the alarm going off happens to be correlated to some degree with the thief breaking into my car. In this case, there is no nomological underpinning of the correlation: the thief doesn't *cause* the alarm to go off. Still, the case satisfies the conditional from above, since there is a correlation between the thief breaking into the car and the alarm going off. The problem with such accidental correlations is that they don't carry over to other situations. If I were to move to a different neighborhood where there's no Friday night car thief, the car alarm would not be helpful in telling whether my car is being broken into. For a representational system to be useful, it must carry correlational information about a non-accidental nomologically connected event, process, or property. So, we need to add as a proviso to the definition of correlational information that the correlation holds non-accidentally, so that it projects to new cases.[7]

Another shortcoming of this definition is that it treats representational states and what they represent as singular states that represent singular states. In most biological representational systems, the system can represent a whole range of states by taking a range of states itself. The thermoceptive system is clearly such a system: there is a range of perceptual states that represent a range of temperature states.

> Correlational information carried by a range of states:
> Let X be a range of states of the representational system and let Y be a range of states of the target of representation.
> a being in a state in X carries correlational information about b being in a state in Y iff
>
> 1. for a nomological reason for every value F of X there is a value G of Y such that $P(Gb|Fa) > P(Gb)$ or;
> 2. for a nomological reason for every value F of X there is a value G of Y such that $P(Gb|Fa) < P(Gb)$.

It needs to be the case that the probability is either lowered or raised *across the range of states*. A system that lowers the probability of its target states in some cases and raises it in others does not carry correlational information.

An early version of a correlational-informational account of mental content was developed by Dretske (1981). Dretske's early was based around the concept of *indication*, which he equates to raising the probability of a certain state obtaining to 1. An event B

[7] Shea adds that this proviso on non-accidentality is bound to a certain spatiotemporal region: a representational state can non-accidentally correlate for some part of the lifespan of an organism (say, only during puberty) or for the entire lifespan, or indeed for parts of the evolutionary history of the organism. The need for non-accidental correlation in order to exploit information is bound to a time and place. (Shea, 2018, p. 77)

indicates an event A if and only if the conditional probability of $A|B = 1$. In many cases, causation will underly indication (B indicates A because B is caused by A), but indication can also be realized if both states have a common cause. The problem with indication as an explanation of representation is that it doesn't allow for *misrepresentation.* A single case of B obtaining without A obtaining means that B does not indicate A. In later work, Dretske adopts the addition of *function* to the definition of representation, in order to deal with the possibility of misrepresentation. (Dretske, 1986)

Carrying correlational information is a very common thing, much more common than representation. Going by this definition, almost any effect in a causal chain carries information about its cause, because effects are produced by their causes according to causal laws, and that an effect obtains usually raises the probability that its cause obtained. It doesn't seem right to say that every mental state is a representation of its causes. Later versions of informational theories appeal to *functions* of mental states, and sometimes to consumer systems – systems that *use* the information carried in mental states. These appeals to function and consumer systems are meant to alleviate problems of liberality (many states are counted as having content) and indeterminacy (the content of a state is indeterminate) faced by causal-informational accounts.

4.5 Informational teleosemantics

When my car alarm goes off, this event carries correlational information about my car being broken into, but it also carries correlational information about my car battery being charged. That the alarm goes off is nomologically dependent on the battery being charged, so this event satisfies the definition above. However, we don't normally think of a car alarm as representing a full car battery. Why not?

The answer is that representations depend on *functions* both to achieve representational status and to determine the content of the representation. The function of the car alarm is to signal that my car is being broken into, and not to signal that the battery is charged. The car alarm going off carries correlational information about multiple events. Which among these events counts as its content depends on its function.

Functions don't only play a role in content determination: that something counts as a representation also depends on it functioning as such. As mentioned, the definition of correlational information is so liberal that every effect carries correlational information about its causes. For some piece of correlational information to count as a representation, it must be produced by a system that has as its function to relay this information to some consumer. A plume of smoke hanging above a forest can be used to find the location of a wildfire because of the correlational information that plumes of smoke carry about wildfires. However, to say the smoke represents the wildfire is wrong, since the wildfire is not a system that has the function of signaling its location through plumes of smoke. Compare the captain of a sinking ship setting off a flare so that the coast guard can locate the ship. In this case, the smoke from the flare is part of a system that has as its function to relay the location of the ship to a consumer (the coast guard). The smoke from the flare is a representation of the location of the ship, but the smoke from the wildfire is not a representation of the location of the fire. In these examples I am appealing to an intuitive notion of *function*, that is likely derived from the intentions of the captain of the ship, or the designer of the car alarm. But functions need not be derived from the intentions of actors in the system, nor from the intentions of designers of the system. Systems can have functions even when they are not designed and don't involve actors with intentions.

Accounts that explain content as depending on carrying information and functions are known as informational teleosemantics. The basic idea is that biological systems have *functions* – something that they *do*. A state in that system can only be counted as a representation if the information carried by the state explains how the system performs its function.

Functions can be understood in a variety of ways. Above, I have used examples of artificial systems. A system like a car alarm has a function in virtue of the purpose it's designed for. A car alarm functions to signal car break-ins because it was designed to do so. If we focus on one state of the car alarm system, we should be able to tell whether it has representational content. Say there's a wire running from a sensor on the driver side window to the central computing unit of the car alarm. If the driver side window breaks, the sensor is activated and an electrical signal travels along this wire. This signal can be said to represent a car break-in, because 1) the presence of the signal correlates with the car being broken into, 2) the function of the system is to report break-ins, 3) the signal carrying correlational information about the break-in explains how the system performs its function: the system in effect implements a simple algorithm that states when such an input signal is received, sound the alarm.

This functional-informational explanation of the car alarm is easy enough, but it appeals to the *design* of the alarm in describing its function. For biological systems, appeal to designer's intentions violates naturalistic tenets, so the notion of function must be given a non-intentional explanation. The common naturalistic explanation that teleosemantics proposes is, that a system's function is what it was selected for. This can be understood in different ways, for example evolutionary phylogenetic selection or feedback reinforcement learning. Below I will give a brief overview of various ways of defining the notion of function at play.

Classic etiological function

The 'classical' view in teleosemantics is that function should be understood as *etiological* function – doing that which it was selected for. Neander offers the following definition:

> Etiological function:
>
> "It is a/the proper function of an item (X) of an organism (O) to do that which items of X's type did to contribute to the inclusive fitness of O's ancestors, and which caused the genotype, of which X is the phenotypic expression, to be selected by

natural selection." (Neander, 1991, p. 174)

This idea is at the heart of teleosemantics itself: that the function of a state is determined by its evolutionary history. The function is that which the trait was selected for. Natural selection is a causal non-intentional process, so etiological function doesn't violate naturalistic tenets. Etiological function in some form or other is part of most teleosemantics theories.

There are a few problems associated with the classic etiological function concept. One is how to think about the functions of *new* traits. Say a mutation occurs which will provide an animal with a distinct reproductive advantage, but that advantage has not occurred yet – does the trait have a function? The etiological function theory would have to deny this, since no selection for this trait has occurred yet. This problem has been put in a forceful form as a *swampman* thought experiment: if by chance a system was created that is molecule-for-molecule identical to some biological system, but lacks evolutionary history, does it have the same functions as the biological system? (Davidson, 1987)

A second, related, problem has to do with vestigial traits and traits that have switched function. A vestigial trait has an evolutionary history that, according to the etiological theory of function, would confer upon it a function. But vestigial traits are, by definition, not currently adaptive – they serve no purpose in the current generation. A similar problem occurs if some trait confers adaptive advantage in a different way now than it did in the organism's ancient evolutionary history. Even if the original function is no longer a current function, the etiological theory will ascribe that ancient function to the current trait, perhaps along with its current function.

Modern history function

An alternative to the classical etiological theory of function emphasizes the *modern* evolutionary history in which a trait is *maintained* rather than the ancient history in which the trait originated.

> Modern history etiological function:
> The function of a trait is its disposition or power which explains the recent maintenance of the trait in a selective context. (Godfrey-Smith, 1994)

The advantage of this theory is that it better handles traits where the recent evolutionary history suggests a different function that the ancient evolutionary history. Godfrey-Smith also argues that it is a more adequate conceptual analysis of some ways the term 'function' is used in biology. The downside is that the power to explain *why a trait is there in the first place* diminishes if we limit the scope of evolutionary history which can inform our function-ascriptions. The modern history function theory also doesn't provide an answer to the problem of *new* traits (such as swamp-traits), since these do not have any evolutionary history, ancient or modern. An option for historical theories of etiological function (ancient or modern) is to simply bite the bullet on new traits and hold that swamp traits have no function until they have been used to some evolutionary advantage.

Besides choosing the modern history version of etiological function over the classical version, perhaps it is possible to *include and prioritize* modern history (and the maintenance of traits under selective pressure) in etiological explanation. I think it is fair to say evolutionary biologists, in ascribing functions to parts or systems of organisms, can appeal to both modern and ancient evolutionary history of an organism.

Forward-looking function

One theory of biological function that explicitly tries to avoid the problem of *new traits* is the forward-looking theory of function put forward by Bigelow and Pargetter (1987).

> Forward looking function:
> "Something has a function just when it confers a survival-enhancing propensity on a creature that possesses it." (Bigelow & Pargetter, 1987, p. 192)

The idea, in line with the propensity theory of *fitness*, is that it is not the actual survival of an organism that bestows function upon a trait, but the propensity (a statistical notion) of that organism to survive in the future, given that it has this trait. This theory deals with the problems faced by historical functions precisely because it doesn't appeal to the history of a trait in ascribing it a function. New traits have functions because they bestow propensities upon the organism.

It is not entirely clear what in Bigelow and Pargetter's view the truthmakers are for these propensities are. On the one hand, they say that ascribing function necessarily depends on referring to future events (p.181), but on the other hand they say this event may never have to occur. They compare survival propensities to the dispositional property of fragility; a vase is fragile even if it never breaks. They further say that this propensity supervenes on the 'morphological characters' of the organism, just like dispositional properties supervene on categorical properties.

The problem with this view lies in fleshing out the details: the function depends on the trait conferring the propensity to survive (or be selected), but it is unclear in what *context* that propensity would exist. Bigelow and Pargetter recognize that a propensity is relativized to a certain environment: the trait confers an advantage only in a certain situation. They stipulate that the function of the trait is dependent on its bestowing a propensity *in the organisms' natural habitat*. Godfrey-Smith argues that specifying the natural habitat of an organism implies a reference to the historical habitat of the species – which brings back problems like those faced by historical accounts of function. A trait can acquire a new function if the organism is placed in a novel habitat. (Godfrey-Smith, 1994, p. 352)

Modal function

A more contemporary proposal that is somewhat related to the forward-looking function account is the idea that the function of a trait is determined by its supporting counterfactual claims. According to this theory, the function of a trait x is to do F if and only if *if x did F*, then it would contribute to the fitness of the organism that has x. (Nanay, 2014)

The subjunctive conditional that does the heavy lifting in this definition is cashed out in terms of Lewis-style possible worlds semantics – although the option of a different semantics for counterfactuals is left open by Nanay. With the use of this apparatus, the modal theory of function can be spelled out:

> Modal function:
> F is a function of organism O's trait x at time t if and only if
> 1. There is a *relatively close* possible world w at which x is doing F at t.
> 2. Doing F at t contributes to O's fitness in w.
> 3. w is closer to the actual world than any world w' in which x does F at t, but this doesn't contribute to O's fitness.

Nanay's modal theory of functions looks to me like a more plausible version of Bigelow and Pargetter's forward-looking functions. The 'propensities' in Bigelow and Pargetter's theory boil down to dispositional properties of traits. A common analysis of dispositional properties is to view them in terms of subjunctive conditionals.[8] The modal analysis given by Nanay could be seen as filling in what a propensity is: the subjunctive claim that a trait would contribute to fitness in relatively close possible worlds.

The problem with Bigelow and Pargetter's account was to spell out the conditions in which the propensities would confer advantage without reference to a historically natural habitat. Nanay's theory limits the scope of the propensities by appealing to 'relatively close' possible worlds. But whether that is successful is a topic of discussion: Neander and Rosenberg argue that for every adaptive trait x that is doing F in the actual world there is a relatively close possible world in which x is doing F*, which confers even more adaptive advantage. Therefore, the modal theory must hold that every trait (that is currently adaptive) is malfunctioning because it *could* function even better. (Neander & Rosenberg, 2012) The criticism trades on the fact that the modal theory of function must spell out the precise meaning of 'relatively close'– this meaning must rule out worlds like the one in which x is doing F*.

The modal function account is importantly motivated by a supposed circularity in other accounts of function. Etiological accounts explain the function of a token trait x by reference to the token traits of the same type that the ancestors of this organism possessed. The function of token x is explained by reference to the history of the trait type. But how are trait types individuated? For example, what is it that makes a heart a heart? The obvious answer is that a heart is something that pumps blood – but this references the *function* of the trait. Nanay maintains that this constitutes a problematic circularity in the ascription of functions: to ascribe a function to a token trait, one must identify the trait type the token belongs to. To identify the tokens belonging to this trait type, one must appeal to the function of those tokens. (Nanay, 2010, 2014) In defense of the etiological function account, it has been

8 There are many counterexamples to the 'simple' subjunctive analysis of dispositional properties, but the basic idea that the essence of a disposition is captured by what the object would do in certain contexts stands as the basis for many theories of dispositions.

claimed that there is no circularity because a function of a trait and the type both supervene on the history of selection – it is selection history that underlies both typing traits and ascribing functions. To properly ascribe a function, you need to know the selection history the trait is a part of, but you don't need to see the trait as any member of a function-type beyond that selection history. (Neander & Rosenberg, 2012)

In my eyes, there is a sense in which the modal account is not a direct competitor of the etiological account: they deal in subtly different kinds of explanation. Niko Tinbergen (1963) recognized that when biologists ask the question why an animal exhibits a certain behavior, there are four types of explanation that can be given.

> 1. In terms of the physiological mechanisms and the physical stimuli that lead to the behavior.
> 2. In terms of the current functions of the behavior.
> 3. In terms of the evolutionary history of the behavior.
> 4. In terms of the development of the behavior in the life of the individual [animal].
> (Quoted from Godfrey-Smith, 1994)

These types of explanations of behavior can also be applied to traits more generally (behavioral or otherwise). The etiological account clearly tries to give an explanation of type (3), while forward-looking and modal accounts of function try to give an explanation of type (2), which avoids the problems with new traits. But this enumeration doesn't list competing explanations, rather it lists complementary explanations.

Surely, on the modal theory, the *actual* causal mechanisms and physical structures of the trait are relevant to the truth of the modal claim about what the trait would do in w. Even though the function ascription is set by the subjunctive conditional, the actual advantage that the trait confers in the real world can be explained in mechanistic type (1) terms as well. The same holds for the evolutionary explanation: the fact that there's a type (3) explanation of the trait in terms of evolutionary history, doesn't mean there is no type (1) explanation of that same trait – the type (3) explanation is the story about how it came to be that this organism has these physiological mechanisms.

The question is whether a modal explanation is the type of explanation that we want to base a theory of function on. One reason to think it perhaps isn't, is that a modal explanation is not a causal explanation. If the success of naturalizing mental content relies on giving an explanation of function in *causal* terms, then perhaps a modal theory of function would mean the project of naturalizing content cannot be successful. But, if a modal explanation is thought to be naturalistic (albeit non-causal), then perhaps there is a chance the project can succeed.[9]

[9] There are fundamental metaphysical issues at stake, and there is no room to discuss them here. I am skeptical about the naturalness of a subjunctive conditional interpretation of dispositions when this is understood as a metaphysical explanation and not a mere semantical explanation. I have defended fundamental dispositional properties in the actual world as the truthmakers of causation in my master's thesis. (van Westen, 2015)

The disagreement between the modal theory and the etiological theory then lies in which of these two available explanations we should see as providing the type of function needed to explain mental content. At this point I don't see a definitive answer to that question: modal function can deal with new traits in a way that etiological function can't, but there are other ways to deny the importance of Swampman thought experiments. (Millikan, 1996; Neander, 1996; Papineau, 2001)

System functions
A rather different notion of function than the etiological notions or the dispositional notions is the one put forward by Cummins (1975). This notion has to do with the (analytical) decomposition of biological systems into sub-systems or parts. The idea is that a complex activity of a system can be realized by the activities of parts of that system. The function of such a part, then, is the contribution it makes to the complex activity of the system. This notion of function is not a naturalizing explanation of function, so it cannot be the whole story about function in a teleosemantic theory. But etiological accounts (or other versions of teleosemantics) may incorporate this idea of functional analysis, and we will see in the next section that the notion of a *task function* has some links to this idea.

4.6 Task functions

The relevant notion of function for teleosemantics, according to Shea (2018), is that of a *task function*. This idea integrates some aspects of the notions of function discussed above, and I will spend a bit more time discussing it.

The main idea of a task function is that systems perform tasks in a way that generates *robust* and *stabilizing* outcomes, and that this constitutes their task function. By robust outcomes it is meant that the system produces the same outcome in response to a range of different inputs and produces it in a range of different external conditions. In the example of the car alarm, a good car alarm produces its output (alarm sounds) no matter whether the break-in happens from the driver side or from the passenger side (different inputs). To a good car alarm, it shouldn't matter if the car is parked curbside or in a parking garage, whether it is raining or sunny, what time of day it is, etc. The alarm should produce the right outcome in response to a break-in in a variety of external situations. If a system S doesn't produce an output F in response to a variety of relevant inputs and under a variety of relevant external circumstances, then it can't be the function of S to produce F.

> <u>Robust outcome function:</u>
> An output F from a system S is a robust outcome function of S iff:
> 1. S produces F in response to a range of different inputs; and
> 2. S produces F in a range of different relevant external conditions.

A stabilized function is a function where the output of the system disposes the system to generate the same output in future cases. A biologically salient example of such a system is an evolutionarily adaptive trait: an output F is generated by a system S because in the organism's evolutionary history outputs like F have contributed to survival or reproduction. A trait is selected for because it increases fitness, which leads to the trait being more common in later generations of that population. For an individual in the later generations, it has that trait *because* it contributed to survival or reproduction for the ancestors of that individual. Here you see how the notion of a task function *includes* etiological functions. Evolution, however, is not the only way in which biological systems can come to stabilize outcomes.

Shea recognizes three main ways in which functions can be stabilized:

1. Through natural selection: an output F is generated by a system S because in the organism's evolutionary history, outputs like F have contributed to reproduction of the trait that produces F.
2. Through persistence of organisms: if S is an organism, an output F is generated by S because it contributes to persistence of S.
3. Through learning with feedback: an output F is generated by a system S because the organism has learned to produce F because of positive feedback.

The first way in which functions can be stabilized is natural selection – which also forms the basis of the etiological account. The precise formulation of this stabilizing function could be tweaked to accommodate the modern history version of etiological function.

The second way functions can be stabilized (stabilizing by contributing to persistence) is a simple mechanism: if something contributes to the organism's survival, then that organism is more likely to survive *with it* than *without it*. Say you are driving, and a truck suddenly swerves into your lane. Your quick reflexes allow you to avoid a collision, saving your life. The output from your reflex system contributes to your survival – which means that your quick reflex system also survives. If you had slow reflexes, things would have ended badly for you and your reflex system. In this way any mechanism that contributes to the persistence of the organism is stabilizing its outcomes.

In learning with feedback, outputs are stabilized because some outputs get positive feedback while others don't (or some don't get negative feedback, while others do). Feedback is a mechanism that increases the likelihood the system will produce a particular outcome in the future.

> Task function:
> An output F from a system S is a task function of S iff:
> a) F is a robust outcome of S
> and
> b) 1. F is a stabilized function of S or
> 2. S has been intentionally designed to produce F

The circularity problem that Nanay attributes to etiological accounts of biological function could also be thought to apply to the task function account, albeit only to the evolutionary stabilizing mechanism. In the case of stabilizing through contributing to survival of the individual organism, there is no reference to a trait type, but rather to the history of that token trait. The same holds for feedback learning: this occurs within the lifespan of the individual organism, and therefore there is no need to reference the trait type – the stabilizing occurs in the history of the token trait.

The problem of new traits persists to some degree for the task function account. As discussed above, 'classic' teleosemantics focuses mostly on the evolutionary way of stabilizing outcomes, which is a comparatively slow process. Relatively recent traits are hard to account for in this theory. Recognizing that the persistence of an individual organism can stabilize outcome functions means that functions can arise within a generation. Feedback learning can also stabilize outputs. So, while task function is still dependent on the history of the organism, the timescale on which functions can develop is much shorter than just the evolutionary timescale. If there are other ways in which outputs can become stabilized, those would also contribute to defining task functions.

Still, the extreme cases of *new* traits, such as swamp traits, are a problem for the task functions account. If an organism has no history whatsoever in which selection or learning occurred, then we cannot ascribe a task function.

The task function account doesn't lend itself to *incorporating* the modal function account: the dispositional properties of a trait are not outcome-stabilizing mechanisms. Rather, the subjunctive conditionals that Nanay appeals to are assertions *that* the organism will produce a particular output under specific conditions. These two types of explanation may be compatible: systems with task functions produce stable and robust outputs, which means they have a certain modal force. But to the task function account, the essence of what makes the system a system that produces outcome F is not the subjunctive conditional that it will do so in certain circumstances, but rather that the outcome-stabilizing mechanism has made the system so.

Task functions are not specific to representations: non-representational systems have functions too. To arrive at a definition of representation we need to combine the idea of task functions with the idea of carrying information.

4.7 Consumers and producers

There is some disagreement in the literature on teleosemantics about how exactly representational content is determined. One camp, known as consumer teleosemantics or benefit-based accounts, holds views like what I have expressed above. The content of a representation is dependent on there being a *consumer system*, which is some further system that *uses* the information held by the representational state. What the consumer system does with the information is what determines the content. *Task functions* are an essentially benefit-based concept: there is a system task and the content of informational states in the system is determined by the role of these states in the successful performance of that task. The stabilization of the *outputs* of the system is of paramount importance in consumer accounts.

On the other side of this debate there is *producer teleosemantics*, which emphasizes the *inputs* of the system which produces the representation. Neander, for example, claims that evolution selects not for a certain output, but for responding to certain inputs. (Neander, 2017)

Take the example of the frog visual system, which is commonly used in discussions of teleosemantics. In the frog's visual system, the firing of retinal ganglion cells is correlated to small black objects in the visual field. The frog uses this correlation to react to bugs, which it can catch with its tongue. The debate between producer and consumer teleosemantics boils down to the question (in this example) of whether the firing of the ganglion cells represents the presence of bugs, or merely the presence of small black objects.

Neander argues that there is no causal link between the firing of the retinal cells of the frog and the nutritional value of the bug. There *is* a causal link on the other hand between the visual properties of the bug (small, black, moving) and the firing of the retinal cells. For Neander, the trait of the frog visual system that evolution has selected for is *that it responds to these visual properties*. Accordingly, she argues that representational content depends on *response functions*, in contrast with Shea's *task functions.* A response function is the function of a system *S* to be caused to go into state *F* in response to an input *I*. On a producer account, the content of that state *F* just is the input *I* that the system is disposed to respond to. Questions of indeterminacy on a producer account are to be answered by determining what exactly the system is causally sensitive to. In the frog example, the firing of the retinal cells is *caused* by objects with certain visual properties, and not by the nutritional value of those objects –the indeterminacy is resolved in favor of the 'small black moving objects' content.

The advantage of the producer-based account is that it doesn't import non-perceptual information into the content of perceptual states. The fact that flies are of nutritional value to a frog is not something a frog can *see.* That this fact plays a role in the content determination on consumer accounts could perhaps be seen as a weakness of these accounts. However, a reverse argument could also be made: the producer account hasn't explained *why* the frog would represent small black moving objects beyond deferring to the brute fact that this confers evolutionary advantage. The consumer-based theory does a better job at explaining what the representation does for the organism, not only what it is produced in response to. In the following, I will be adopting a consumer-based variant of teleosemantics (using task functions rather than response functions) precisely for this reason.

4.8 Representation based on correlation

The notion of correlational information, combined with the idea of a task function can give us an account of representations. The idea is that a system produces representations if some component of that system carries correlational information about some condition, and this correlational information *explains* how the system performs its task function.

> Representational content:
> If component R of a system S with task function F_j carries unmediated explanatory

information about condition C, then R represents C. (Shea, 2018)

There are some terms in this definition that need to be unpacked. *Unmediated explanatory information* in this definition means that the correlation between R and C must be what primarily explains the success of S in achieving its task function.

In the frog fly-catching example, the firing of retinal ganglion cells is correlated to small black objects in the visual field. This correlation can figure in an explanation of how this output is stabilized, but it can only do so if there is an extra element in the explanation that states that *small black object* correlates with *edible bug* in the environment of the frog. For why would reacting to small black objects confer an adaptive advantage if *small black objects* didn't correlate with *edible bugs*. Without this second correlation we'd be left wondering how the firing cells contribute to any evolutionary advantage. The correlation between retinal ganglion cells firing and the small black object only figures in a *mediated* explanation, while an *unmediated* explanation is available: the correlation between the retinal ganglion cells firing and the presence of an edible bug.

> Unmediated explanatory information:
> The unmediated explanatory information carried by a set of components R_i in a system S with task functions F_j is the exploitable correlational information carried by the R_i which plays an unmediated role in explaining, through the R_i implementing an algorithm, S's performance of task functions F_j.

The implementation of an algorithm mentioned in the definition refers to the computations that are applied to the R_i in order for the system to arrive at its output. For example, the frog system may have some set of rules to determine when to stick its tongue out based on the firing of the ganglion cells, plus perhaps some other conditions.

The clause that the correlational information a representation uses be *unmediated* helps alleviate issues of indeterminacy in content. If there is more than one correlation that could explain the functioning of the system, the one that does so without the need for additional correlations is the one that we should take to be the content.[10]

To pull back from this technical talk for a moment: task functions are just a way of saying *what a system is doing*. What the frog visual system is doing *for the frog* is better described as recognizing bugs than as recognizing small black objects. Why? Because bugs contribute to the frog's survival and the visual recognition of bugs gives a frog an adaptive advantage. This is an important point that I will be applying in the case of thermoception later.

---

[10] Determining the content in cases in which the candidate contents are co-extensive in the environment of the animal are a bit more difficult. Say that a frog lives in an environment where all black moving objects are edible bugs. We might still insist that what explains the stabilized outcome is the nutritional value of bugs, not the visual properties of the black moving object. Cases where the candidate contents are necessarily co-extensive (or some other strong modal form of co-extensivity) are even more difficult. I would argue that such indeterminacy of content is not problematic, because the content is only indeterminate on a level of description that outstrips the individuation of relevant elements of the animal's environment.

To summarize, representations are those components in a system that help the system achieve its task function *because* they carry unmediated explanatory correlational information. Correlational information *per se* does not constitute representation, it is the exploitation of correlational information by some system that achieves robust and stabilizing outcomes within a certain task by using that information that makes a component of that system a representation.

4.9 Beyond 'simple' correlation

The theory of representation sketched above is a very simple form of representation: the system exploits a correlational relation between its representational states and the thing it is representing. What remains underdeveloped in this view is the *informational* properties of this relation that go beyond simple correlation. The underlying idea of how correlations carry information is that they reduce uncertainty: a correlation makes it the case that if you observe the signal, that raises the chance that the state it signals occurred.

Information theory, as originally developed by Shannon (1948) offers ways to quantify the amount of information carried and shows how system properties influence the amount of information carried. Correlations are a quantification of reduction of uncertainty, but other system aspects play into the equation: a signal is transmitted through a channel, and real-world information channels have limited capacities, called the *rate*. Real informational systems also have *noise* and *distortion*: the signal received on the other end of the channel is not necessarily the same as the signal sent. In the case of distortion, the change in the signal between sender and receiver is systematic i.e., the same sender signal always results in the same receiver signal. In the case of noise, the change the signal undergoes is not always the same, so it can be represented as a stochastic process.

Most results in information theory concern how information can be faithfully transmitted over a limited and noisy channel. Martinez (2019) argues that philosophical accounts of representation do themselves a disservice by not taking into account the finer points of information theory. Biological perceptual systems are exactly such systems that need to confer faithful information about states with huge entropy (the complicated world) through noisy and limited channels.

Martinez argues that representations occur in information processing systems that function at rate-distortion sweet spots. In general, a narrower channel (low rate) causes more distortion. Given a particular pair of an original message (what is represented) and decoded message, a rate-distortion function gives a minimum rate at which a given distortion is possible. The curve that represents this rate-distortion ratio can have sweet spots: increasing the rate from such a sweet spot achieves relatively little reduction of distortion.

Martinez's argument goes beyond claiming that rate-distortion sweet spots are where representations *occur*, he actually *defines* a representation as a signal in a representational system that has an encoder-decoder strategy that occupies a sweet spot in the rate-distortion curve.

How do these systems come to have such rate-distortion sweet spots? One answer is that this happens when systems aim at representing natural kinds: especially if one thinks of natural kinds as instantiating *homeostatic property clusters.*[11] Natural clusters within the field of property instantiations are what begets rate-distortion sweet spots. And, not incidentally, they are also the kind of thing organisms have a use for representing. Systems that produce *robust and stabilized outcomes* are exactly such systems that respond to natural clusters and, if Martinez is right, they are systems that realize rate-distortion sweet spots.

Now there are two ways one might endorse the idea that rate-distortion sweet spots are important to representations. The first is that the existence of such a sweet spot, and the system using it, are constitutive of something to be a representation. That seems to be the line taken by Martinez in giving a definition of a representation:

> **The Rate-Distortion Approach:** A signal, S, in a certain information-processing pipeline, P, is a representation if the following two conditions are met: **Existence:** There are sweet spots in the rate-distortion curve associated with P. **Optimality:** S is produced as part of an encoder-decoder strategy that occupies the vicinity of one of these sweet spots. (Martínez, 2019)

The second way to endorse the idea that rate-distortion sweet spots are important to representations is to take on board that this is what representation *tend to do*, and to take on board that it explains in part how systems achieve robust outcomes, but to deny that representations are exclusively defined by it. This is the approach I would advocate.

Defining representations exclusively by such sweet spots misses an important part of what biological representations are: they are not just informational entities with certain information-theoretic properties, they are biological entities with biological functions. Animals don't represent environmental properties just because they happen to be clustered, they represent properties because that allows them to successfully achieve tasks with stabilized outcomes. It's not clear to me that Martinez would disagree with this idea, because he believes "that teleofunctions have a role to play in a complete theory of representation" (2019), but to me it seems that *function* isn't just an add-on to informational theory of representation, but a constitutive element of a representation. A further reason to include functions is that, as we've seen above, functions aid in giving a theory of representational *content.* The definition of representation given by Martinez's rate-distortion approach seems to answer the *status question* but leaves out a satisfying answer to the content question. The role that teleosemantics can play relative to the rate-distortion approach may well be answering the content question.

Another approach to representation that goes beyond 'mere' correlation is an account that characterizes representations as mirroring the structure of what they represent. I will go into this theory in the next chapter.

---

[11] Martinez seems to endorse a theory that natural kinds simply are homeostatic property clusters, but it is enough for the argument that natural kinds instantiate homeostatic property clusters.

5 Exploiting structural representations

A central idea in the literature on perceptual representations is that mental states or processes can represent properties or processes because they share structural properties. This idea comes in various guises: as *analog representation*, as *structural correspondence*, as *similarity* and as *isomorphism* or *homomorphism*. These ways of spelling out the same essential idea are not equivalent, and I will say something about each of them.

My view is that structural representations are a necessary *addition* to a theory of representation based on carrying information if one wants to explain certain representational abilities of organisms. In this chapter I develop an account of structural representation that is weaker than claims of isomorphism and homomorphism, and therefore more realistic. I show that such a weak account is enough to explain some behavioral abilities associated with structural representation.

In section 5.1, I introduce the concept of structural representation via the roundabout way of the similarity account of representation. Section 5.2 discusses the relation between structural representation and *analog* representation. That notion is related, but due to the divergent ways of defining analog representation, we need to separate that notion clearly from structural representation. Section 5.3 goes into the representational abilities of structural representation. What can a structural representational system do that a mere correlational system cannot? Sections 5.4 through 5.7 are about the minimum 'technical' requirements for structural representation. These sections work towards a definition of *structural correspondence:* the relation that underlies structural representation. Some authors have said structural representations depend on isomorphisms (or homomorphisms) being realized between representational states and what they represent. I take this claim and investigate whether structural representation could occur when there is no such isomorphism. I conclude that weaker structural links can be enough to give a representational system some abilities of structural representation. This is what I call structural correspondence. Section 5.8 discusses a proposal for a structuralist account of mental representation that is in some respects similar to my account, but which I ultimately reject.

5.1 The similarity account of representation

One way in which a system might represent is by being *similar* to what it represents. This might be somewhat plausible as a theory of pictorial representation: a picture of a daisy represents a daisy because it looks similar. As a theory of mental representation, a simple similarity account doesn't really hold water and isn't an option anyone has cared to defend, as far as I know. However, thinking about the shortcomings of a similarity account can help us gain insight into some aspects of structural representation.

<u>Naïve similarity:</u>
M represents T iff M and T are similar.

The first problem with this account is that similarity and representation have different logical properties. Similarity is a reflexive and symmetric relation, but representation is not. (Goodman, 1976) The second problem is that 'similarity' is a rather unconstrained term: what counts as similar or not is vague. Any two things can be considered similar in some way. (Goodman, 1972, p. 443) If *naïve similarity* is amended with a clause that says M must be similar to T *in a relevant way*, then this notion of relevance is doing the heavy lifting. Still, similarity is too inclusive: there are many cases in which two things are similar, but neither represents the other. Putnam (1981, p. 1) imagines an ant that traces a line in the sand which happens to look like a caricature of Winston Churchill. Although the line in the sand bears a certain similarity to Churchill, the ant has no knowledge of Churchill nor any intent of representing Churchill. To say the line in the sand represents Churchill is stretching the meaning of the term. The third problem is that similarity can't be a necessary condition for representation: the word 'Churchill' represents a former Prime Minister of England but bears no (relevant) similarity to the man.

There are further problems with similarity accounts of representation: it has trouble accounting for misrepresentation (Suárez, 2003) and it has trouble explaining how we can represent non-existing things. Some problems may be eliminated by appealing to the intentions or beliefs of the people using these representations. While that could possibly work for pictures or scientific models, it is a non-starter for a general theory of mental representation which needs to also explain the representational status of beliefs and intentions.

So, as a general theory of representation, we can discard the similarity account. But there is a certain truth to the intuition that drove the account. Some representational systems allow users to exploit the similarity between the representation and what's represented. That we can use a map to find our way in unfamiliar surroundings depends on a certain correspondence of features of that map with features of the environment. For example, if cities A and C are further apart (in reality) than cities A and B, then they should be further apart on the map. The map allows us to make judgements of relative distance because the relative distance of points on the map corresponds to the relative distance of the cities that are represented.[1]

We may be tempted to think now that this type of representation works because the map *shares properties* with reality. But this needn't be the case. Two-dimensional maps can use color to indicate altitude, e.g., darker colors for higher altitudes and lighter colors for lower altitudes. If the color scale used has the right kind of structure, then we can make comparative judgements about the altitude of places that are represented on the maps. But these places don't actually have those colors. We are using one range of properties (colors)

---

[1] Not considering for the moment the distortion inherent in projecting a curved surface (the world) onto a flat surface (the map).

to represent another range of properties (altitudes). Such representational schemes are extremely common and very useful. For example, a timeline allows us to visually represent points in time, even though temporal location is not something that can normally be *seen.*

In mental representation, such representational schemes may also be used. If we think of the firing rate of certain sense receptors as representing the intensity of a particular stimulus, then this looks like such a representational scheme: one range of properties (stimulus intensities) is represented by another (firing rates).

This way of representing depends on *structural correspondence* of what is being represented. What that means exactly is explained in this chapter. Before diving into a theory of structural representation, I will discuss the related but not identical notion of *analog representation.*

5.2 Analog representations

Representation is sometimes classified as *analog* or *digital*. Typically, those are taken to be mutually exclusive. There are several accounts of what exactly it means for representation to be analog or digital. On one traditional view analog representations are *continuous* and digital representations are *discrete.* On a different view, representations are analog if they mirror the structure of what they represent.

Continuous/discrete

According to Goodman (1976), a representational system is analog only if it is *dense* and *non-differentiated*, while digital systems are necessarily *discontinuous* and *differentiated.*[2] Take as an example a system that uses marks to represent numbers: in a digital system, it is (theoretically) possible to determine which marks are copies of one another – i.e., which marks represent the same number. This is why a digital system is *discontinuous*. It is also possible in a digital system to determine exactly which number a given mark represents. This is why a digital system is *differentiated.*

In an analog system, for any two marks that are not copies, there is an intermediate mark that is not a copy of either (density). And for any two marks that represent different numbers, there is an intermediate number represented by a mark which is not a copy of either (non-differentiated). A familiar way of putting these definitions is that analog systems use continuous scales to measure continuous variables, while digital systems have discrete representations that stand for discrete variables. On this view, the difference between analog and digital information depends on the structure of the representations themselves, and the structure of what they represent. Systems that represent continuous variables along continuous dimensions are analog, while digital representations represent discrete variable using discrete dimensions. An example of an analog representation would be a mercury thermometer. The height of the mercury column varies continually with temperature: for every two values on the representational dimension, there is another value that lies between

[2] See Schonbein (2014) for a more contemporary presentation of this view.

them. Digital representations on the received view are discrete: between two points on the dimension lies a finite number of other points.

Against Goodman's view it has been argued that analog computers can have differentiated and non-dense representations. Think of a computer that takes values of resistance as inputs – this can be realized either with a continuously variable resistor (potentiometer) or with a rotary switch (with a finite number of positions) that has individual resistors of different values attached to each switch position. The rotary switch differentiates the inputs and makes them non-dense, but that doesn't change the fact that the computer is an analog computer.[3] That analog systems don't use discrete and differentiated representations is not necessarily true. (Lewis, 1971; Maley, 2011)

Mirroring

An alternative view on the analog/digital distinction holds that representations are analog computers if they are *structurally isomorphic* to or otherwise *mirror* what they represent. On this view analog representations need not imply continuous or dense. A discrete structure can be represented in an analog way. One such view is Maley's, who argues that a representation is analog if the represented quantity covaries with the representational medium. (Maley, 2011) A similar view is presented by Kulvicki, who argues that analog representations are essentially structure-preserving, and that this explains the unique representational capacities of analog systems (especially continuous ones). (Kulvicki, 2015) In cognitive science, a similar view has been presented to explain how mental rotations are computed. (Shepard, 1978)

This notion of analog representation as preserving structure is essentially identical to what I call *structural representation.* But, to avoid confusion with the notion of analog content as essentially continuous, and possibly other notions of analogicity, I will use the term *structural representation.* As I explicate that notion in the remainder of the chapter, you may hold in mind that this is essentially also an explanation of the minimal requirements of the mirroring conception of analog representation.

Analog magnitude representations

There is considerable literature in cognitive science on Analog Magnitude Representations (AMRs): representational systems that represent magnitudes in an analog way. The notion of analogicity used here is the mirroring view: they are in essence what I call *structural representations:* the representational abilities of the system depend on the representations preserving structure that is present in the representational targets. (Beck, 2015)

Structural representations are not uncommon in the animal kingdom. The most commonly discussed AMRs represent spatial magnitudes (i.e., distances), time, or quantity. (See for example: Dehaene, Piazza, Pinel, & Cohen, 2003; Gallistel, 1993; Walsh, 2003).

Peacocke (2019, Chapter 2) offers an insightful discussion of the relation between the *metaphysics* of a magnitude (what I would call the *structure)* and analog representations of

---

[3] Haugeland (1981) argues that the computer that takes input from individual resistor values is digital, but see Maley (2011) for a defense and expansion of Lewis' argument.

that magnitude. His discussion applies to representationally extensive magnitudes: those magnitudes for which a calculus of addition and ratios holds, and for which we have representations that can also be added, multiplied, etc. This isn't the case for temperature: (meta)physically speaking 20°C is not twice as warm as 10°C, and it is also not perceived as such. Analog extensive representations of extensive magnitudes (such as spatial magnitudes) preserve structure to a greater degree than is required by my definition of structural representation. They are structural representations of a very robust kind.

Although I chose to use the term 'structural representation' rather than 'analog magnitude representation' I take my account put forward in this chapter to hold true for AMRs: what makes these representational systems analog magnitude representations is the way they preserve structure – a notion I will explicate in the following sections. One could well argue that a thermoceptive system with structural representations *is* an analog magnitude system: it represents the magnitude of temperature in an analog way.

5.3 What can structural representations do?

A structural representation is when a system represents relations between its representational targets by means of a relation between its representational vehicles.

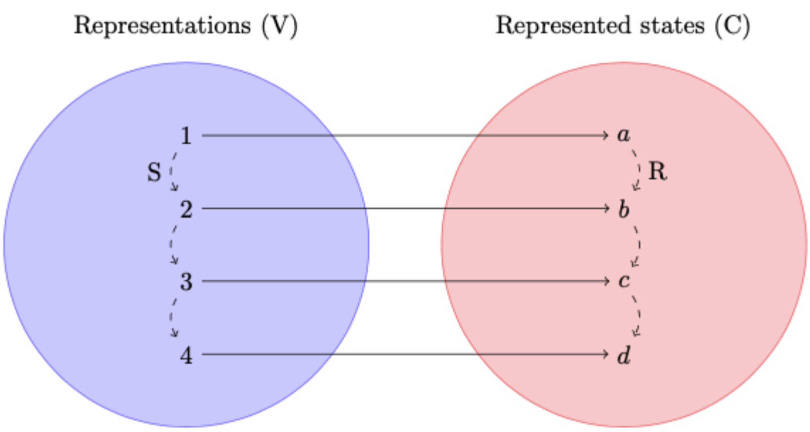


Figure 5: a diagram of a structural representation. Relation S over representational vehicles mirrors relation R over represented states.

In figure 5 we see a set of representations (vehicles) that are mapped to represented states (contents). There is a relation S over the representational vehicles such that it *mirrors* the relation R over the states or objects that are represented. I will say more about what precisely *mirror relations* are, but the above example is one.

Structural representations endow a representational system with special representational capacities. Below I explain the most important capacities of structural representations that set these systems apart from merely correlative systems.

Ordering

If the relation R is a (partial) ordering relation, then S can provide the organism with information about the ordering of represented states on R. Think of a mercury thermometer: that 20°C is *warmer than* 10°C can be read of the thermometer. This is because the relation *higher column of mercury* preserves the relation *warmer than*, and that relation provides an ordering of temperatures.[4] A mere correlational representational system doesn't provide this. Preserving ordering relations allows organisms to represent quantities: if the relation '*bigger than*' (or a similar relation, such as '*hotter than'*) is preserved by S, then organisms can perceive that some represented state is bigger than some other, because the vehicles that represent those states stand in the relation S.

More generally, mirror relations allow for processing that depends on the formal properties of the relation represented, not only ordering. For example, if R is transitive, its mirror relation S should be transitive as well. An organism equipped with such a structural representational system could use the transitivity of S in processing about R. When structure is preserved, the type of computations that could be made about R can be made about S, so the organism can use S to compute things about R.

Addition

Some representational systems preserve structure in such a way that it allows for addition. For example, representations of *distance* can have this property. Think of any good ruler: the distance between the mark of 0 and 5cm represents (or can be used to represent) a distance of 5cm. The distance between the 5cm mark and the 6cm mark represents a distance of 1cm. The representational system here allows addition: if we add the distance between the 0cm and 5cm marks to the distance between the 5cm and 6cm marks, we get a represented distance of 6cm. This is also true for the actual magnitude of distance: a 5cm distance plus a 1cm distance is a 6cm distance.[5]

Ratios

Some relations have structure such that it allows for comparing intervals. For example, real numbers exhibit such structure that we can not only say things like '5 is bigger than 2' but

---

4 The example of a thermometer (or an analog fuel gauge) is sometimes used as an example of something that carries information without representing. The idea there is that a thermometer cannot misrepresent – the height of the column of mercury is directly causally linked to the temperature of the mercury. Even if the thermometer was not properly calibrated, and the numbers that indicate temperatures on the side of the glass were way off, the height of the column of mercury would still covary reliably with temperature. It is important to note here that correlation (and the preservation of structure) is, on my view, not sufficient for representation. It is only when a downstream consumer (e.g., a doctor) uses the information carried by the thermometer to perform some task that the possibility of misrepresentation appears. That preservation of structure is neither sufficient nor necessary for representation will be explained further later I this chapter.

5 This toy example is a bit obvious since it uses distance to represent distance in a 1:1 ratio. O'Brien and Opie call this first-order resemblance. But we can expand the example: a technical drawing uses distance to represent distance but in a ratio of, say, 20:1. And mental representations of distance may allow addition, while distance in the world is not represented by distance in the brain. This would be an example of what O'Brien and Opie call second order resemblance.

also ‘the interval between 5 and 2 is greater than the interval between 0.11 and 3.21 or we can say that 12 is twice as large as 6. Real numbers allow a calculus of ratios, which brings into play mathematical operations such as multiplication and division.

Some representational systems allow for a calculus of ratios and represent magnitudes that allow for such a calculus. Think again of the ruler: the distance between the 0cm and 10cm mark stands in a 1:2 ratio to the distance between the 0cm and 5cm mark.

Respresentational systems that allow for additive computation and a calculus of ratios over magnitudes that also allow for additive computation and a calculus of rations are what Peacocke (2019) calls *extensive magnitude representations.* Temperature perception does not fit this description: perception of temperature doesn’t allow for addition or a calculus of ratios. It is not true that 20°C feels twice as warm as 10°C. It isn’t even true from a (meta)physical point of view that 20°C is twice as warm as 10°C.

Preserving structure plays a role in animal cognition and perception, such as in spatial representations. Preserving the metric structure of spatial relations allows organisms to compare distances between points in space.

An example of spatial structure being preserved is in place cells in the rat hippocampus. Place cell activation correlates with location – so that when the rat is in a certain location withing a rectangular enclosure, certain cells fire. This realizes a correlational system of representation: the firing of place cells correlates to the rat being in a certain location. (O’ Keefe & Burgess, 1996) The location of the place cells in the rat hippocampus is not *spatially* correlated to the locations they respond to – the hippocampus does not form a topographic map. However, when a place cell is activated, other place cells that correlate to nearby location are co-activated. This co-activation realizes an important mirror relation: the *nearby* relation on places is mirrored by the co-activation relation on place cells.

When at sleep or resting, the place cells are activated in an offline manner, that plays a role in spatial memory. The rats *replay* routes they took previously. There is also evidence that rats use the hippocampal ‘map’ not only to remember routes, but also to *plan* routes. (Pfeiffer & Foster, 2013) Place cell routes can be observed to be activated *before* the rat takes that route, in so-called pre-play.

It is possible that rats use this kind of structural representation to make computations about distances: the rat could, for example, pre-play several routes to the same point and choose the route that took the shortest amount of time to pre-play (replay length is proportional to route length). What we can see from this example is that certain preservations of structure may give an organism the ability to compare distances or intervals in physical quantities, because the structural properties of those quantities are to some extent preserved in relations on the representational vehicles.

The cases where structure is preserved to such an extent to allow a calculus of addition and ratios over a domain are some of the most robust representational systems.

Fixing reference of new representational states

Shea (2018) argues that an important feature of structural representation is that when a new vehicle is introduced in a system of structural representations then its relations to other representational states in the system can fix its place in the representational system. That new vehicle comes to represent some state that corresponds to that place in the system. For example, if we introduced a new place cell into the rat hippocampus, with co-activation to some cells but not others, then that cell would come to represent a place in the enclosure due to its connection to other place cells – the location it represents being defined by the pattern of co-activation with other place cells. This way that new representational states can be added to the system allows for sensory systems to be responsive to novel stimuli.

To add to Shea here, the new cell would still need to be responsive to the rat being in a certain place for it to be a *proper* place cell. If the place cell is not activated by any specific place, but only co-activated by other place cells, it looks like we've introduced a 'phantom' location to the cognitive map – a representation of a place that is close to other places but doesn't correspond to a *real* location. Moreover, if the new place cell were to be co-activated with cells that correspond to distant locations (e.g., on opposite sides of the enclosure) then we'll have introduced an *impossible* location in the cognitive map.

If, on the other hand, we were to introduce a place cell that fires when the rat is in a particular location but doesn't get co-activated by other place cells and doesn't cause co-activation of other place cells, this would still be a *real* place cell since it would (through correlation) represent an actual location in the rat's enclosure. However, this isolated place cell would be of no use in navigation tasks or in remembering routes since it has no pattern of co-activation to other place cells.

So, while I agree with Shea that 'new' place cells can gain representational power by taking a certain position in the existing network, I hold that the underlying correlation with an actual location is a precondition for it being a place cell in the first place. This is true more generally: structural representations tend to work *because* the individual elements of the representational system carry correlational information about the environment. Reference fixing, in this case, is done by the correlation, not the structure. But perhaps we can say there is something like *structure fixing*: the process by which vehicles come to represent structural properties of the environment. Structure fixing is done by the vehicle taking a place in a structure of vehicles, such as the pattern of co-activation in rat place cells.

This is an important thing to note, since it shows that a system of structural representation does not take the place of a correlational system of representation, but rather is an addition to it.

Fixing reference for the entire system

In some representational systems it may be structure *and only* structure that explains how the system represents. In such a case, the reference is fixed not by correlation, but by structural representation. Think for example of a map of a symmetrical space. Say there is a labyrinth that has two entrances: one in the west and one in the east. The sequence of left and right turns one needs to make to reach the center is the same whether one enters from the east or the west. If the map has some marking of direction or origin (e.g., a compass rose, or a label that marks one entrance as east or west), and you know which entrance you are taking, then the lines on the map correlate with the walls of the labyrinth. However, if the map were unmarked, such that either entrance of the map could refer to the east or west entrance, then there is no exploitable correlation between points on the map and locations in the labyrinth, because each point on the map could equally well represent a location in the west half of the labyrinth as in the east half. However, since the sequence of left and right turns is the same when entering from the east or from the west, the map can still be used to effectively navigate to the center of the labyrinth. The effectiveness of the map in this task is explained solely by the structure of the map corresponding to the structure of the labyrinth, and not by correlation between individual points in the labyrinth and individual points on the map. This is, however, a contrived case. For sensory systems, it looks like some type of correlational representational scheme underlies structural representations.

5.4 Isomorphism

Some authors have claimed that the preservation of structure is a central feature of sensory representation. Some might even say it is a requirement for representation: no representation without preservation of structure. This view seems false to me, as correlational accounts of representation do not require preservation of structure. However, it may be true that structural representation is a common or even ubiquitous feature of sensory systems.

The preservation of structure that's needed for structural representation has sometimes been characterized as an isomorphism:

> The "correspondence" that a system producing intentional representations is designed to establish between these representations and their represented can be thought of as an abstract isomorphism, in this way. Transformations (in the abstract mathematical sense) of the representations correspond to transformations of what is represented, such that different representations map different represented in a systematic or "productive" way. (Millikan, 2006a, p. 198)

Or from the field of cognitive psychology:

> The brain is said to represent an aspect of the environment when there is a functioning isomorphism between an aspect of the environment and a brain process that adapts the animal's behavior to it. (Gallistel, 1993, p.18)

Authors who have put forward similar theses, that structural representation realized by isomorphisms is the backbone of perceptual representation include Cummins (1996), O'Brien (1998), Millikan (2006b), Gallistel & King (2009), and Churchland (2012). Akins considers this to be "the traditional picture of the senses" (1996)

It must be noted here that it is not usually claimed by the proponents of the isomorphism requirement that this is a sufficient condition for representation. Most theories of representation hold there are other additional requirements.[6]

---

[6] Pure similarity accounts of representation are perhaps an exception to this rule, but as argued in the beginning of this chapter, there are strong counterarguments against this position.

How do structural representations work? We've seen some examples above, both abstract and concrete, but these leave underexplored the precise requirements that the structural correspondence needs to satisfy for structural representation to be possible. What are the formal requirements associated with the loose notions of structural correspondence? Is it the mathematical notion of an isomorphism as some authors mentioned above suggest?

Actual properties of parts of the environment and actual representations are not mathematical objects. But the structural features of properties in the environment and of representations can be abstracted as mathematical objects. An apparatus of sets, functions and structures will provide us with a way to evaluate whether a sensory system satisfies the isomorphism requirement.

Let C be the set of represented temperatures states. Let V be the set of temperature representations. What we need to say whether C represents V, is some specific sort of mapping of the members of V to the members of C. This can be expressed as a function $f: V \rightarrow C$. The function can be defined as a triple (V, C, F), where F is a functional subset of the cartesian product $V \times C$.

What this means for temperature perception, in non-mathematical terms, is that there is a set of temperatures (the set C), and a set of perceptions (the set V), and a way these are paired up (the pairs are given by the function f). Then we can look at what qualities this function has; is it injective, is it surjective or maybe both? Does the mapping preserve certain structures?[7]

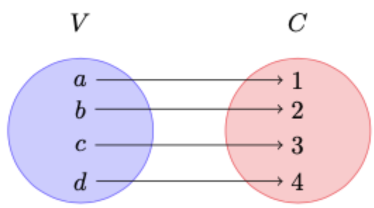


Figure 6: A mapping (represented by arrows) from V to C

Isomorphism is a mathematical property of mappings, and it has somewhat different applications in different fields of mathematics. Other related terms from mathematics are sometimes used in the literature on representation: *homomorphism, morphism, homeomorphism,* etc. In most of the philosophical literature the terminology is not used in a very rigorous manner. All these terms have to do with the properties of mappings, and the preservation of structure. One requirement that is essential to the notion of isomorphism is that the mapping from V to C is both *injective* and *surjective*, also known as *bijective.* So minimally, an isomorphism is a bijective mapping. An injective mapping is a mapping where

[7] The definition of functions used here, and the notions of injective, surjective and bijective mappings, and structures are all due to (Bourbaki, 2004).

each member of the domain V is mapped to a unique member of the codomain C. [8] A surjective mapping is a mapping such that for each element y in the codomain C there is at least one element x in the domain V, such that $f(x)$=y. What this means is that each element in C has an element of V that is mapped onto it. [9]

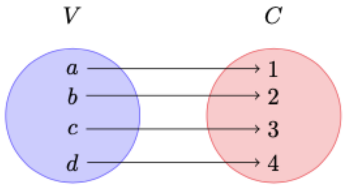
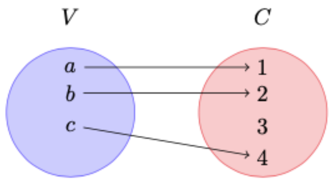
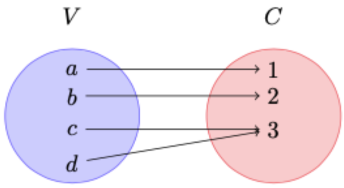
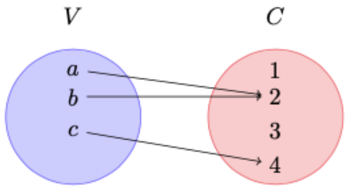
	Surjective	Non-surjective
Injective	 (Bijective)	
Non-injective		

Table 1: properties of mapping functions.

Injection

You can see how a bijective mapping between represented and representation is good for representations: it guarantees that each content in the representational scheme has a vehicle assigned to it (surjective, and that one vehicle always corresponds to only one content (injective). This is why such a mapping is sometimes called a one-to-one correspondence.

Now, depending on the branch of mathematics involved, the definition for isomorphism can include other requirements that are to do with the preservation of structure. I will deal with structures below. But first, let's take a closer look at the requirement that the mapping between represented and representation be injective and surjective.

So, what speaks for counting injectiveness as a requirement for a representational relation, is that it gets rid of equivocation from two vehicles to a single content. Whether such a requirement is likely to be satisfied by a sensory system depends on how vehicles and contents are individuated. If we individuate vehicles by their representational contents, then injectiveness is always satisfied. If we individuate vehicles in some other way, then I may be the case that two vehicles have the same content, so that injectiveness is not satisfied. It is likely that actual sensory systems sometimes have redundant representations: multiple representational vehicles with the same content.

[8] Formally: $f: V \rightarrow C$ is injective if $\forall x, y \in V, f(x) = f(y) \rightarrow x = y$.

[9] Formally: $f: V \rightarrow C$ is surjective if $\forall y \in C, \exists x \in V: f(x) = y$

Surjection

What speaks for surjectiveness as a criterium for representation is that it stipulates that all the contents in a representational scheme must be linked to some vehicle. Whether this makes sense as a requirement for structural representation depends on how the set of contents is defined. Take thermoception. If C contains all possible temperatures, then human thermoception clearly is not surjective, since we do not have perceptual states corresponding to *all possible temperatures.* The range of the human thermoceptive system is limited. But if we specify that C contains those temperatures that are in the representational scheme (i.e., those that have arrows going to them in a diagram like the one above), then surjectiveness is automatically satisfied. So, we are left wondering why surjection would be a requirement for structural representation: on a narrow reading it is *always* satisfied, and on a wide reading it is likely never satisfied.

Inverse isomorphism

It could also be a requirement that an isomorphism has an inverse – that there is a function f^i such that $\forall x \in V,\ \forall y \in C,\ f(x) = y \rightarrow f^i(y) = x$. Meaning that for each arrow from vehicles to contents, there is a return arrow from that content to the corresponding vehicle. Then, if that function f^i is also injective and surjective, this has consequences for the demands on the representational scheme.

Inverse injectiveness (i.e., injectiveness of the inverse mapping function) would mean that each vehicle can only have one content: there are no representations that correspond to more than one content. The consequence is that such a representational system wouldn't equivocate different contents – it would be able to distinguish each content. This is a good requirement for a representational scheme, but it is also unrealistic for sensory systems. Perceptual systems tend not to be as fine-grained as the range of possible states they respond to. There is normally some degree of equivocation, so as a strict requirement this seems unhelpful.

Inverse surjectiveness gets rid of superfluous representations. If f^i is surjective, then there can be no representations that are not mapped to a content. This demand would mean that a representational scheme would not contain any representations that don't refer. This seems to be a feasible requirement, but also not very necessary to spell out. A mental state that does not have any connection to temperatures would not be likely to be counted as part of a representational system for temperatures.

Continuous, Dense, Discrete

The images in this chapter display toy examples using finite sets for V and C. In mental representation of magnitudes, it's possible that the set of possible contents is infinite: it may, for example, be true that for every two distinct distances, there is an intermediate distance that is longer than the shorter distance and shorter than the longer distance. Biological representational systems may have a finite number of distinct states. If this is the case, then for representations of infinite sets either surjectiveness is not satisfied and there are magnitudes (contents) that are not represented in the system – or inverse injectiveness is not satisfied and there are vehicles that correspond to more than one magnitude.

The issue recalls Goodman's definition of analogicity. If the representational system is not dense and continuous, then it cannot be isomorphic and inversely isomorphic to a set of objects that *is* dense and continuous.

Perhaps it could be an argument against the isomorphism view that it is impossible that perceptual systems have an infinite number of states they can take – they can't be continuous and dense because animals are finite creatures. This would be a mistaken argument. The finitude of animals does not mean that representational systems cannot make use of infinities: for example, the firing rate of a single neuron is potentially a continuous and dense dimension: it could be true that for every two firing rates there is an intermediate firing rate.[10]

A second argument against this point is that a discrete system can still be a case of analog representation. (Lewis, 1971; Maley, 2011) On the mirroring conception of analogicity there is no requirement for the representational system to be continuous or dense.

So, while the toy examples may seem to apply only to discrete finite sets, in fact they could apply also to dense, continuous, infinite sets. In the written definitions there is nothing that implies either V or C needs to be finite.

5.5 Order

Temperature can be thought of as an ordered set. What I meant by this is the intuitive idea that there is a natural order to temperature: they can be arranged in a low-to-high fashion. The temperature of a cold day in February in New York is colder than on a hot day in August in Mississippi. My coffee was hotter five minutes ago than it is now, etc. This is not just a consequence of the units we use to measure temperature (e.g., degrees Kelvin or Celsius), but a feature of the physical magnitude.

This order can be described as a binary relation on the set of temperatures. This relation is transitive, antisymmetric, and semi-connex.

> <u>Transitivity:</u> if $a \leq b$, and $b \leq c$, then $a \leq c$.[11]
> <u>Antisymmetry:</u> if $a \leq b$, and a is not b, then b can't be lower than or equal to a.
> <u>Semi connex:</u> if a and b are in T, then either a is lower than b or b is lower than a.

We can see that the relation of *smaller than or equal to* imposes a (partial) order on the set of temperatures.

It could be a requirement that representations need to *preserve the order* of the represented. But it is important to note that even when C and V have a similar ordering

[10] Here's another (highly speculative) counterargument against this criticism: the universe itself may be discrete, as physicists working on quantum gravity tend to think. (Dowker, 2006) If that is the case, then the set C will always be finite.

[11] Formally: a relation R is transitive if $\forall a, b, c \in X: (aRb \land bRc) \rightarrow aRc$
a relation R is antisymmetric if $\forall x, y \in X,\ (xRy \land yRx) \rightarrow x = y$
a relation R is semi connex if $\forall x, y\ (x \in X \land y \in X \land x \neq y) \rightarrow (xRy \lor yRx)$

relation, that does not mean that $f(x)$ has the same place in the order on C as the place x has in the order on V. Let me illustrate:

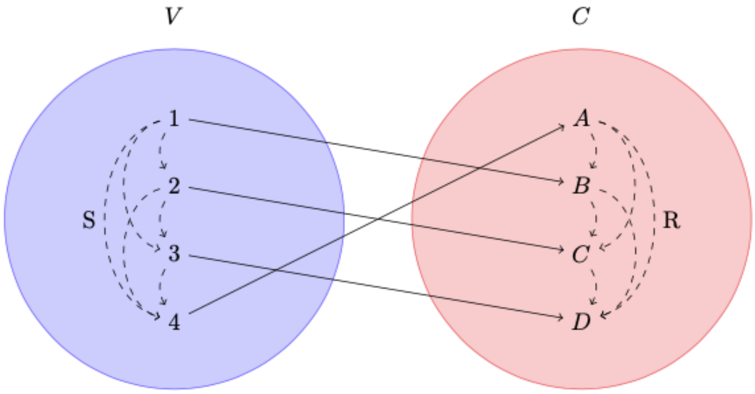


Figure 7: A bijective mapping between two sets that both have the same (partially ordered) structure. But the first element in S is not mapped to the first element in R.

In this example there is a bijective mapping, and both sets have an ordering relation with identical properties. Yet this mapping would not do for a structural representation because the position of the representations in the structure S is not the same as the position of the corresponding elements in structure R. That both structures are ordering structures is not enough for structural representation. What we are looking for when we say representation should preserve order or structure, is not just that there be similar relations in the domain and codomain, but something stronger. In figure 7 the relation S is structurally similar to R, but it isn't a *mirror relation* of R. What that means is discussed in the next section.

5.6 Mirror relations

When I perceive one bowl of soup to be hotter than the other, I seem to be comparing two temperature representations, and making a comparison about the relative temperatures of two objects in my environment. The process could be something like this:

1. I have temperature perceptions E_A and E_B of soup A and B
2. Temperature perception E_A stands to E_B in a certain relation S
3. The actual temperatures of A and B stand in a relation R *hotter than.*
4. S structurally represents R
5. I correctly perceive that soup A is hotter than soup B on the basis of S(E_A, E_B).

If we think a structural representational system should be able to justify a reasoning like 1-5 above, then we should think that merely having a similar ordering relation is not enough. That two elements in the domain stand in a certain relation is no guarantee that their corresponding *values* in the codomain stand in that relation to one another. What we need is for relations to have a *mirror relation* in the codomain.[12] A relation has a mirror relation

[12] A term borrowed from Palmer (1999, p. 77).

if every pair[13] in the relation is mapped to a pair in the codomain, that stands under a different relation. S is a mirror of R if each pair $R(x, y)$ is mapped to pair $S(f(x), f(y))$.
On the example given in figure 7 it is not true that $\forall x \forall y \in V: (xSy) \rightarrow f(x)Rf(y)$. This means that the S in figure 7 is not a mirror of R. Here's the full definition of a mirror relation:

> Let there be a set V and a set C, and a mapping function $f: V \rightarrow C$. Let R be a relation over C, and S be a relation over V. S is a mirror of R iff:
> $\forall x, y \in V, (xSy) \rightarrow \exists a, b \in T: (aRb \wedge f(x) = a \wedge f(y) = b)$

Note that it isn't just the structure of the relation S that makes it a mirror of R. Whether or not S is a mirror of R depends entirely on the mapping f. Here's an example of a mirror relation:

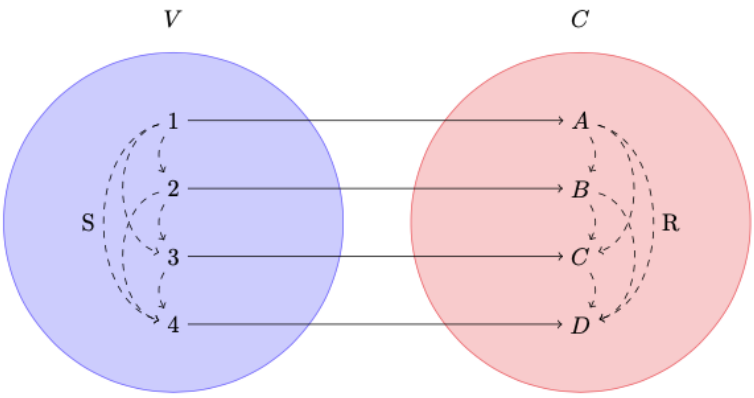


Figure 8: S is a mirror relation of R.

What is interesting about this preservation of structure is that it gives us the kind of information we look for in structural representations: the structure on vehicles allows the organism to respond to the structure on the represented objects.

An important point of difference between figure 7 and figure 8 is that not all bijective mappings preserve all structure. In figure 7, the mapping f does not preserve the structure R, and S is not a mirror relation of R. It is of course possible to come up with a mapping on these two sets such that the relation *is* preserved. You can do it by specifying that: $F\{(1, A)(2, D)(3, C)(4, B)\}$. Mathematically speaking, it is easy for there to *exist* a structure-preserving isomorphism between these two sets. But in the real-world examples of perceptual representation, we don't get to pick and choose our mappings. They are determined by the correlational account of representation spelled out in the previous chapter. Recall the problem with the similarity account of representation, that it is too liberal: similarity is more common than representation. In structural representation, the liberality is constrained by the mapping function: not every structure on vehicles that is similar to a structure on objects is a structural representation. It depends on the mapping function whether some structure can be a structural representation.

---

[13] The notion of a mirror relation can easily be extended from binary relations to relations of higher arity.

Preservation of mirror relations is an important part of representation. It captures a part of what we think representation should be like. When theorists say there should be isomorphism between represented and representation, the charitable interpretation of this statement is to take it to mean that the representations should have mirror relations of the relations of the represented.

In fig. 8 I gave an example of a *bijective* mapping that preserves a relation in a mirror relation. But note that the definition of a mirror relation does not imply that the mapping is bijective. In fact, it is possible for mirror relations to be preserved on mappings that are non-injective and non-surjective.

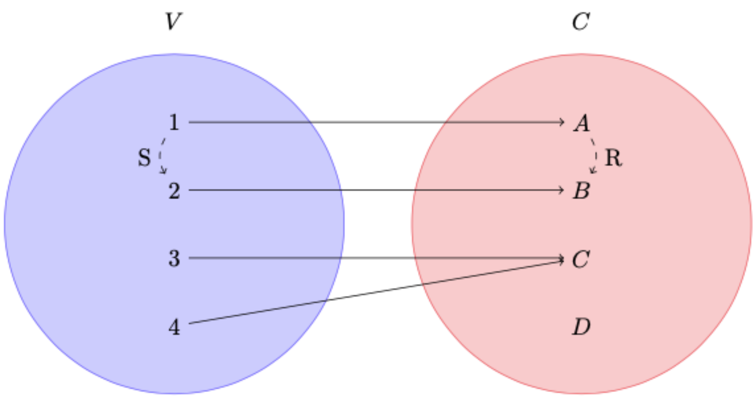


Figure 9: A mirror relation on a mapping that is neither injective nor surjective.

In figure 9 we can see that it is not necessary that a mapping be bijective for a relation to be preserved in a mirror relation. But do note that if we look only at the elements that are in relations, we have a bijective mapping between them. There is a subfunction of f that *is* an isomorphism. If we were to define a function f' that maps the same as f but only takes relata of R as input, we would get a fully structure preserving bijective mapping.

Figure 9 gives us some hope that perceptual systems can indeed use mirror relations to provide structural representations. If it is true that sensory systems routinely fail to have a bijective mapping from inputs to sensory states, perhaps they can have bijective sub-mappings that do preserve structure.

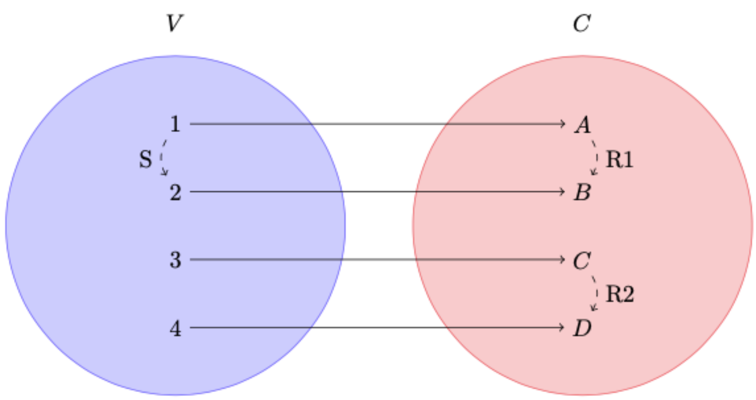


Figure 10: This mapping lets S preserve R1 as a mirror, but there's no mirror to preserve R2. The mapping also happens to be bijective.

Figure 10 presents another logical possibility: a mapping that preserves some structure, but not all of it. Just like in figure 9, there is a bijective relation between the members of the preserved relation and the members of the mirror relation. If it is a requirement for representation that it preserves *all* structure on a domain, then figure 10 fails this requirement. But the mapping in figure 10 *does* preserve the structure $R1$, and that might be enough for the purpose of structurally representing $R1$. This could also be interesting to the case of thermoception, and perceptual representations in general: it is possible to selectively preserve structure.

Now compare figures 8, 9, and 10. The mapping in figure 8 is supremely suitable for representational purposes because it preserves *all* relations, and it does so with no equivocation and with no superfluous representation. The mapping of temperatures to temperature experience for an *ideal* thermoceptive system would look something like figure 8.

The mapping in figure 9 is not as good for representation as the one in figure 8. This is neither injective nor surjective. As discussed earlier, these are disadvantages for representational systems, but it is likely that our senses suffer some of these disadvantages. What *is* good about the mapping in figure 9 is that it fully preserves the structure in a mirror relation. What I think figure 9 shows is that a requirement for structural representation should allow for non-bijectiveness.

Figure 10 shows that it is possible to preserve some structure, but not all. I think it is likely that most sensory systems are like this. While 10 is a bijective mapping, that need not be. You can imagine an example that combines figure 9 and figure 10, that preserves some structure, but not all, and is not injective nor surjective. A mapping like this could not strictly be called *isomorphic,* but it might still carry over enough information to be called some kind of representation. Perhaps we should say that the mapping in figure 10 is representational *qua* the relations it preserves, but not representational of the domain as a whole. This would be a weaker kind of representation.

5.7 Structural correspondence

What we are interested in when we investigate whether V represents C, is if the relations between elements of C can be read off from the relations of elements of V. One interpretation, then, of the claim that representations should be isomorphic, is that the term is used loosely, and what is actually meant is that representation should preserve structure in mirror relations.[14] A more literal interpretation is that these theorists actually take it to be the case that representational mappings should be bijective and preserve all structure. This claim, according to the arguments above, seems wrong. A more charitable interpretation is that for structural representation *of a particular relation* to be possible, there needs to be an isomorphic sub-mapping realized in the representational system.

[14] Neander (2017) shares this interpretation, as do some of the other authors mentioned in section 5.4.

Bijection plus full preservation of structure is a rather demanding requirement of representational systems, and not all elements of that requirement are very informative of what constitutes representation. Rather, the preservation of structure is what counts, and that does not imply isomorphism. We are left having to define what preservation of structure is if it doesn't isomorphism. I call this *structural correspondence.*

We can define a structural correspondence as an ordered triple $\{f, R, S\}$ of a mapping function f, a relation R on the domain of f and a relation S on the codomain of f. For $\{f, R, S\}$ to be a structural correspondence, there must be a part of the function that realizes a bijective mapping. This partial function relates a part of the extension of R to a relation S in a structure preserving manner.

> <u>Structural Correspondence:</u>
> Let f be a mapping function $f: V \to C$. Let R be a structure on C and S be a structure on V. The ordered triple $\{f, R, S\}$ is a structural correspondence iff:
> There is a bijective function $f': A \to B$ such that
> Condition 1: $A \subseteq V \ \wedge \ B \subseteq C$
> Condition 2: $\forall x \in A, \ f'(x) = f(x)$
> Condition 3: $\forall x, y \ \in A, (xRy) \to \exists a\,, b \in B: (aSb \ \wedge f'(x) = a \ \wedge f'(y) = b)$

Condition 1 says that the domain and codomain of f' should be parts of the domain and codomain of f. Condition 2 says that the mapping f' produces for its elements is the same as the mapping f produces for those elements. Condition 3 is the condition for mirror relations but applied only to the domain and codomain of f'.

This definition allows for non-injective and non-surjective mapping functions. It says nothing about whether $f: V \to C$ is structure preserving. What it does require is that some *part* of the function f is a bijective structure preserving mapping.

It is a big advantage of this definition that it does not require that the entirety of structure R be preserved in S. It needs only preserve the structure R on those elements of C that are in B. What that means is that the definition allows representations to preserve part of a structure. Think of a thermometer a doctor might use to take your temperature: it preserves the ordering structure of temperature within a certain range of temperatures, but it does not preserve the ordering structure of the entire range of possible temperatures. So, the mapping of all thermometer states to all temperature states does not preserve the ordering, but the sub-mapping for the range in which the thermometer functions properly does preserve the structure of that segment of the temperature range.

This lines up with some of the more sophisticated statements of the isomorphism requirement that are found in the literature. For example:

> Ordinary maps preserve the relative positions of points, coded by, for example, Cartesian coordinates. Hence they preserve every kind of geometrical relation because geometrical relations such as "perpendicular to" or "collinear with" exist among points or point sets (lines and surfaces) by virtue of their relative positions. However, it is possible to have maps that do not preserve relative positions in the full

> sense of the term. These "weaker" maps preserve only certain classes of geometric relations. (Gallistel, 1993)

A similar sentiment is expressed here:

> More formally, a representational system can be analyzed as a homomorphism: a mapping from objects in one domain (the world) to objects in another domain (the internal world) such that relations among objects in the external world are mirrored by corresponding relations among corresponding objects in the representation. (Palmer, 1999, p. 77)

Gallistel's quotation reflects my conclusion that not *all* structure needs to be preserved for something to be interesting as a representation. Palmer's quote does suggest (through lack of quantification) that *all* structure needs to be preserved – I read him as saying all relations should have mirror relations – but he allows that the morphism not be bijective. He uses the term *homomorphism* to signal that representational mappings don't need to be bijective and structure preserving, as on the isomorphism requirement.

My definition of *structural correspondence* formalizes the intuition that representations to some extent preserve structures of what they represent. Unlike the isomorphism requirement, it does so without being overly demanding, because it does not require bijectiveness, and does not require *all* structure to be preserved. Because it is less demanding, it is more likely that actual sensory systems satisfy my requirement than the isomorphism requirement.

Representational systems can realize structural correspondence but still not be structural representations. Structural correspondence is a necessary but not sufficient requirement for structural representation. For a representational system that exhibits structural correspondence to be system of structural representations, it needs to put that structural correspondence to representational use. If the relation of the vehicles is of no consequence to downstream processing or the behavior of the organism, then it is not doing any representational work. This is essentially the same as with correlational representation: the structure on the vehicles needs to be an ineliminable part of the explanation of how the system performs its task function. (Shea, 2018)

Structural correspondence comes in degrees: some representational systems may preserve more structure than others (and exploit said preservation of structure). The extensive magnitude representations that Peacocke discusses are high on that scale: they preserve structure to such an extent that a calculus of addition and ratios over representational vehicles preserves a similar calculus over the magnitudes they represent. A system like that has strong representational abilities that can be especially useful in e.g., spatial navigation. The thermoceptive system has much more limited preservation of structure and consequently has more limited representational capacities.

5.8 A structuralist theory of mental representation?

My definition of Structural Correspondence is essentially identical to the definition that O'Brien and Opie (2004) give for what they call the 'weak form' of *second-order resemblance*. The 'strong form' of second-order resemblance implies isomorphism, while the weak form allows for more limited mirror relations, just like *structural correspondence.* O'Brien and Opie argue that isomorphism is not a necessary requirement for structural representation, and that indeed the weak form of second-order resemblance is the apt analysis of structural representation. This part of their view I share. However, the *structuralist* theory of mental representation that O'Brien and Opie develop form the notion of second-order resemblance is quite different from the line that I take, which includes informational teleosemantics.

O'Brien and Opie think that one of the main desiderata for a theory of mental content is that it explains how mental states *cause* the organism to have an appropriate response to the environment. I agree. But they also claim that, because of this desideratum, the definition of the content fixing relation cannot contain any reference to something like *adequate responses* or a *well-functioning system.* For if mental content is fixed by some relation between vehicles and objects that appeals to the adequate response of the organism to the environment, then it cannot be a causal explanation of how that organism adequately responds to the environment, due to circularity.

Because of this worry, the structuralist account of representation says that the content of mental states should be fixed without reference to *functions* and the like, if those are to be understood as *causal roles*. Rather, content is fixed by the second order resemblance relation: the content of a state is what it resembles.

I think this view is problematic for two related reasons: first – it is unclear how to determine what a given system resembles: the discovery of some second order resemblance in a representational system seems to depend on the assumption that that system represents a certain set of objects. Second, second order resemblance is a very weak requirement – almost everything bears some degree of second order resemblance to almost anything else. The same is true for Structural Correspondence, but on my account that is not the *whole* story of mental representation.

To expand on my first issue with the structuralist view, here's a quote from O'Brien and Opie:

> Suppose $S_V = (V, R_V)$ is a system comprising a set V of objects, and a set R_V of relations defined over the members of V. […] We will say that there is a relation of *second-order resemblance* between two systems $S_V = (V, R_V)$ and $S_O = (O, R_O)$ if, for at least *some* objects in V and *some* relations in R_V there is a one-to-one mapping from R_V to R_O such that when a relation in R_V holds of objects in V, the **corresponding** relation in R_O holds of the **corresponding** objects in O. (O'Brien & Opie, 2004) (bold type my emphasis)

The problem with this definition is that there is no prior notion of *corresponding relation* or *corresponding object* available to the structuralist. If second-order resemblance is supposed to be what fixes the reference of a mental representation, then the definition cannot contain pairs of vehicles (in set V) and objects (in set O) for which it is *given* that this vehicle represents that object. And the same holds for the relations R_V and R_O: how can it be determined, using this definition, whether second-order resemblance holds unless we already know which relations on set R are the counterparts of which relations on set V?

Since I said that the definition of second-order resemblance is essentially identical to my definitions of structural correspondence given above, it would seem that the same issue would arise for my own account. But this is not the case: in my account the mapping function *f* determines which members of V represent which members of O, and that in turn helps determine which relations on O are represented by which relations on V (since relations are individuated by their extensions).

The second issue with the structuralist theory is this: if we take second-order resemblance to be the content fixing relation, then any system that bears second order resemblance to some other system carries that other system as content. The issue here is that second order resemblance is very liberal. The definition specifies that there must be a one-to-one mapping for at least *some* elements in V and *some* relations R. If we take 'some' to mean 'at least one', then what's minimally required to show second-order resemblance between V and O is a one-to-one mapping of one element of V to one element of O, and a one-to-one mapping of one relation defined on that element of V to one relation defined on that element of O. This is easily done: a mapping from a single element to a single element is always a one-to-one mapping. And for the relation we can always choose the relation of identity: since every object is identical to itself, there will always be a mapping of R_V to R_O for that one element in either set. Therefore, any set V has a massive number of sets that it resembles. It seems to me that O'Brien and Opie would either need to say there is something more to content fixing than just second-order resemblance or admit that any set V has a massive amount of content. Essentially Goodman's criticism that everything is similar to everything else to some extent still holds for the more sophisticated proposal of second-order resemblance.

That leaves one more point from the structuralist account to contend with: even if I am not convinced that the structuralist account succeeds at fixing content, does their argument about the circularity inherent in appealing to *functions* pose a problem for my account? I think it doesn't. The notion of task function that I adopted (from Shea) in chapter 4 does not make explicit reference to something like an organism's disposition to adequately react to the environment. Rather, it looks at robust and stabilizing outcomes that can be the product of a variety of historical causal processes.

6 A representational theory of thermoception

This chapter presents my view of thermoception as a sense modality that represents temperatures. In this chapter I integrate the ideas presented in the previous chapters. I conceive of thermoception as a sense modality, as argued in chapters 1 and 2. My model of thermoception essentially incorporates the naïve temperature representation model presented in chapter 3 and adds onto it the more sophisticated theory of representation presented in chapters 4 and 5.

I use this framework to explain phenomena related to thermoception that can seem surprising on a naïve view. Such phenomena can be explained as a *function* of the system or must be counted as a systematic malfunction. In the background of this project there is an important role being played by Akins' idea that senses are narcissistic and that the central question to ask is *what a sense is doing*. I see the notion of a *task function* as essentially capturing this idea and transforming it into a way of determining the intentional content of perceptions. The project is to explain both 'normal' functioning and some surprising phenomena as part of the performance of the thermoceptive system's task functions.

In section 6.1 I introduce the three types of thermal objects that the thermoceptive system attributes temperatures to and that constitute the three modes of temperature perception. In 6.2 I discuss the first of these modes, the attribution of core body temperatures. 6.3 elaborates on the second mode, which attributes temperatures to parts of the body. 6.4 deals with the mode that attributes temperatures to external objects. In section 6.5 I discuss Akins's theory of 'narcissistic' senses, which is critical of a representational model of thermoception. In 6.6 I dive deeper into problems with a representational account, specifically problems that challenge a correlational view of representation for thermocept. In 6.7 structural representations of temperature are under consideration. 6.8 discusses a relatively recent finding in the neurobiology of temperature perception, which suggests that pathways for automatic thermoregulation and behavioral thermoregulation may be quite distinct. I will discuss the possible consequences for my view. In section 6.9 I discuss several temperature illusions: surprising phenomena that may be difficult to explain from a representational point of view.

6.1 Three modes of temperature representation

Thermoception is a sense modality that represents temperatures of the body, parts of the body, or of objects external to the body with the goal of adapting behavior and thermoregulation to the circumstances in the environment.

Let's unpack this statement a bit. Firstly, I call thermoception a sense modality. I have argued that claim in chapter 2, so I will not repeat those arguments here. The second part of the definition recognizes three kinds of objects of thermoception: temperatures of the body (i.e., core temperature), temperatures of parts of the body, and temperatures of objects external to the body.

Thermal sensations can be categorized as having three different kinds of thermal objects: representing core body temperature, representing body-part (skin) temperature, or representing object temperatures.

1. Attributing core temperature. Thermoception contributes to central thermoregulation in a manner that is not area specific. The signal contributes to the thermoregulatory system by signaling low temperature for example, but its contribution is to the core temperature system. This is typically unconscious.
2. Representing the temperature of a part of the body (i.e., the skin on the hand) to be a certain temperature (e.g., feeling your hand to be cold). This may be conscious or unconscious.
3. Object temperature: necessarily connected to touch. Often conscious. This is the part of thermoception that is most like canonical cases of perception in that it is exteroceptive.

These three thermal objects aren't really objects in the sense of material objects. They are perceptual objects: the kinds of quantities the perceptual system is attuned to. Each type of thermal object can be seen as a different way of functioning of the thermoceptive system. It perhaps makes more sense to speak of three *modes* of the thermoceptive system, with each mode attributing temperature to different objects: to ourselves (core temperature), to body parts, or to external objects.

The word 'modes' may take you back to the dual modes from the Armstrong/Vesey debate and the Haptic/Tactile distinction made by Matthen. This is not coincidence. In the haptic mode of thermoception, the perceptual object is an area of the skin – so the haptic mode of thermoception is equivalent to attributing thermal properties to an area of the skin. The tactile mode of thermoception attributes thermal properties to external objects – so it is connected to the third type of thermal object in the list above.

6.2 Core temperature representations

The type of thermoceptive activity that contributes to physiological thermoregulation is perhaps not usually thought of as perception. It tends not to enter into consciousness and is essentially interoceptive. I have included it here as a type of thermoception because of the continuity between the thermoreceptor activity that contributes to central physiological thermoregulation and the activity that contributes to conscious temperature experiences: it is the same thermoreceptors activating in response to the same types of stimuli – it's the processing that makes the difference.

The consensus view of thermoregulation is that of a dynamic collection of thermoeffector circuits (e.g., for shivering or sweating) that are independently driven by a unique combination of core temperature signals and ‘shell’ temperature signals.[1] Core temperature is much more influential in driving thermoeffectors – it provides the main (negative) feedback for thermoregulation. Shell temperatures provide positive or negative auxiliary feedback. (Romanovsky, 2018)

As I said above, the temperature signals that contribute to thermoregulation are usually unconscious – our bodies are thermoregulating all the time, but we are not constantly conscious of our core temperature. That doesn’t mean it is impossible for us to consciously perceive our core temperature. Especially in cases where the core temperature is high or low, it may become conscious.

The shell temperature signals (coming from the skin) that provide inputs to the physiological thermoregulatory system tend not to become conscious individually: when you are cold and shivering, you don’t tend to feel it as a sensation in specific areas of the body but rather as an overall coldness. Of course, you may feel cold overall and at the same time feel some specific area to be exceptionally cold, e.g., the fingers, but the overall sensation of coldness is not located in specific body parts. What becomes conscious is the weighted body temperature to which receptors throughout the body contribute. The same thermoreceptor responses happening in the fingers in such a case may be contributing to central physiological thermoregulation as well as to the conscious sensation of your fingers being cold. But not all the areas in which receptors are providing input to the thermoregulatory will feel specifically cold.

For some thermoeffectors the spatial location of the active receptor may matter: when cold receptors in the palm of the hand are activated, they may cause vasodilation at that location. But this is not always the case: for *central* thermoeffectors it may not matter whether the shell signal is coming from the hand or the foot.

Plausibly, a lot of the time when there are thermoeffectors active, you won’t be consciously feeling cold or hot.[2] Even when you do feel cold or hot, the variable that becomes conscious is usually a weighted average with little spatial content.

Even though the process described above is not very close to the canonical picture of a sense, I think it certainly counts as sensory activity by the criteria presented in previous chapters. It uses appropriately wired-up sense receptors to provide the organism with temperature information on which behavior (as well as physiological responses) can be based. Animal behavior doesn’t just respond to sensations of cold paws or warm rocks, it very importantly responds to low or high body temperature.

The account of representation put forward in the previous chapters also seems to fit well with this functioning of the thermoceptive system. The thermoregulatory system is a consumer of signals generated by thermoreceptors, and it uses the (thermal) information carried by those signals to implement an algorithm that produces a stable output under a variety of circumstances. In fact: stabilizing core temperature *is* the main function of the thermoregulatory system. As discussed in chapter 4, for a state to be a representation,

1 Thermoeffectors are processes by which an organism regulates its body temperature.
2 This is a somewhat unsubstantiated claim: perhaps some people are hyper-aware of their body temperature.

function is of importance. For these sensory states that provide the inputs to the thermoregulatory system, providing temperature information to this system is their biological function. Whether you endorse the etiological or another account of function, it seems straightforward that reliable inputs of the thermoregulatory system are of huge adaptive importance to an organism that can only survive within a narrow range of body temperatures.

The sensory process that produces body temperature representations is different from a canonical sense modality in that it is *interoceptive*. Body temperature representations are interoceptive not because they make use of thermoreceptors *inside* the body – receptors inside the body can be used to glean information about outside stimuli. The retina is *inside* the head, yet vision is the canonical example of exteroception. What makes body temperature representations interoceptive is that they don't attribute thermal properties to any external objects, but only to the body itself. This attribution is what essentially differentiates the modes 1-3 listed above, not the location of the receptors that are involved.

6.3 Located skin temperature representations

What I call located skin temperature representations here should be distinguished from the sleeve thermoreceptor signals that feed into the body temperature variable. Rather, I am talking about representations of specific parts of the body being a certain temperature. You can feel your foot to be cold without feeling cold overall – in such a case you are having a conscious representation of your foot as being a certain (low) temperature.

Although I just used an example of a *conscious* representation of the temperature of an area of skin, these types of representations need not always be conscious. You might unconsciously be wiggling your toes in your shoes because they are cold, or you might retract your leg while sitting by a fire because it was getting a little hot, without becoming aware of it.

These thermal objects, compared to *core temperature,* have much more explicit spatial content. The essential difference between this mode of thermoception and the body temperature mode is that the perceptual object that properties are attributed to is a spatially individuated *area* of the body. If the body moves, the stimulus moves with it.

As mentioned in the previous paragraph, local thermal stimuli can play a role in physiological thermoregulation – local thermoeffectors can be triggered by local thermal stimuli. Located skin temperature representations also can drive thermoregulatory behavior, such as adapting your posture or putting on gloves.

Located skin temperature representations, like core temperature representations, are interoceptive because they attribute temperatures to (parts of) the body rather than to external objects. However, they perfectly fit the modified neuroethological account as they are representations of a physical quantity that allow the organism to employ certain types of behavior.

The distinction between core temperature representations and located skin temperature representations can in some cases be vague. As said, shell temperatures can contribute auxiliary signals to core temperature regulation, and likely shell temperatures also contribute to conscious perception of core temperature. So, what is the principled difference between core temperature representations and located skin temperature representations? My claim is that it is a difference in spatial attribution. Located skin temperature sensation are attributed to specific of the body: they have spatial content in that they allow you to say that a certain *part* is feeling cold while another is not. This content need not be very determinate: the boundaries of the cold area may be hard to identify precisely. If the spatial content is very indeterminate, it may even be hard to clearly distinguish between core temperature representations and located temperature sensation with poor spatial content. But in essence, I think core temperature representations are different from located temperature representations in that they simply have no significant spatial content. It is not the case that when you feel cold, it feels like the coldness stops at the edge of your body. There is no three-dimensional body-shaped spatial content to a core temperature representation. In located skin temperature representations there is spatial content, even if it is indeterminate.

6.4 Object temperature

The mode of thermoception which attributes temperatures to objects is what in chapter 2 I have called the *haptic mode* of thermoception. It is the type of temperature sensation where a thermal stimulus on a particular area of the body is bound to a tactile (touch) stimulus and both which results in attributing the temperature to an external object. This mode of temperature perception is multimodal: it involves touch.

The spatial content of object temperature perception is allocentric. Since it attributes thermal properties to external objects that are touched, it locates these stimuli in an object-relative space, not in an egocentric space.

Object temperature perception contrasts with core temperature perception and skin temperature perception in that it is clearly exteroceptive because it attributes properties to external objects rather than to parts of the body. Object temperature attribution is probably more commonly conscious than skin temperature attribution and core temperature attribution. This ties in with the specific function of object temperature attribution in thermoregulation. For physiological thermoregulation it doesn't matter whether a stimulus is attributed to an object or to the skin – the system should react in the same way. But for behavioral thermoregulation it does matter. If a temperature stimulus is located in allocentric space rather than in egocentric space, motor behavior aimed at *avoiding* stimuli makes more sense. It makes sense to try to avoid a hot stove, but not to try to avoid a hot hand. Object temperature perception can lead to more sophisticated and diverse behavioral responses than core temperature or skin temperature perception. It makes sense for this type of perception to become conscious if adequate responses are on a more sophisticated behavioral level rather than physiological responses or automatic behavioral responses.

Core temperature	Skin temperature	Object temperature
limited spatial content	Egocentric spatial content	Allocentric spatial content
interoceptive	interoceptive	exteroceptive
Unimodal	Unimodal	Multimodal (touch)
More often unconscious	Unconscious or conscious	More often conscious
Goal: thermoregulation	Goal: behavioral thermoregulation	Goal: thermal behavior

Table 2: three modes of temperature perception.

Can there be attribution of temperature to external objects in the absence of touch? I think this may be possible in the case of radiant heat. We can attribute temperature to a hot grill without touching it, based on the heat it radiates. Movement plays an important role in the spatial attribution here: the origin (and to some extent, size and shape) of the radiation is inferred from the areas of skin that are warmed and the change of that area as we move our body relative to the source of radiation.[3] The spatial content in cases of radiant heat will usually be less determinate than in cases that involve touch; touch gives us sharp boundaries for material objects, whereas radiant heat perception (without the help of other senses) can give only a relatively vague shape, size, and location. I find it likely that it depends on the determinacy of this spatial information whether radiant heat is attributed to an external object in allocentric space, or whether it only contributes to located skin temperature perception. My hypothesis would be that in cases of more determinate spatial content, there can be object temperature perception without touch.[4] This conjecture could be an interesting subject for further empirical investigation.

Aside from the case of radiant heat, there can be cases of indirect touch that give rise to object temperature experiences. Touching a warm (or cold) object through a medium (say, a thin nitril glove like used in medical contexts) can be sufficiently like directly touching an object to cause an object temperature attribution. In fact, all temperature and touch perceptions of external objects are mediated in some sense since mechanoreceptors and thermoreceptors are located in subsurface layers of the skin.

Gray (2023) argues against the idea that distal temperature perception is possible through indirect touch. He contrasts it with haptic touch, where a medium can provide information about the shape or texture of a distal object in such a way that we feel the shape and texture of that object itself; think of a walking stick that allows us to feel the shape of an object we explore with the tip, or think of how, while driving, we can feel the surface of the road

---

[3] Gray (2023) denies that a location can be inferred without assumptions about the intensity of the radiant heat source. But this is only true if our exploratory movement is limited to one dimension: closer to or further from the source. If we can move around the heat source, then it can be located in space without assumptions about its intensity. Whether this really leads to attribution of temperatures to an external object would probably depend on the specifics of the situation.

[4] Contra Matthen (2021) who holds that on its own, temperature perception cannot give rise to the haptic mode of touch.

through the vibrations of a moving car. The ability to perceive the properties of distal objects depends on the medium: Gray notes that we can't perceive texture or shape of distal objects through a slack rope, because "the intermediary must be able to convey information about the tactual properties of distal objects".

Gray claims that there is no equivalent phenomenon in temperature perception: a metal rod conducts heat, but probing hot objects with a metal rod does not lead to an experience of that distal object as hot, but rather to an experience of the metal rod as hot. In a more debatable example, he also claims that "when you touch something hot with gloves on, it is your gloves that tend to feel hot not what is touched". In a footnote he concedes that this depends on the type of glove: non-insulating gloves may allow for distal object perception. But Gray thinks this is not a relevant case, because non insulating gloves "constitute a minimal thermal barrier" and this scenario "would be more like not wearing gloves at all".

I find this argument confusing. In the case of haptic touch, the criterion for an adequate medium is said to be that it transfers information about the tactile properties of the distal object. But in the thermal case, Gray seems to disallow media that transfer thermal information very well because they are a "minimal thermal barrier". I would say that a rigid stick, in the haptic case, is a 'minimal vibrational barrier', and that therefore it allows us to perceive vibration in distal objects. The rigidity of the stick vs. the flexibility of the slack rope plays the same role in the haptic case that the thin gloves vs. the thick gloves play in the thermal case, so they should be treated equally. The criterion that the medium must be capable of transferring information should be applied in the thermal case as it is in the haptic case.

To return to the metal rod example: doesn't that show that the perception of thermal properties of distal objects is impossible even with a thermally transparent medium? I don't think so. Besides being thermally conductive, a metal rod also has significant heat capacity. That means that it will take a relatively long time before the metal rod has assumed the temperature of the distal object. And when we retract the rod, it will only gradually cool down, remaining warm for a while after we stop touching the distal object. This property of the medium interferes with the transfer of thermal information because it limits our ability to actively thermally explore the object. We can compare probing thermal distal objects with a metal rod to exploring tactile distal objects with a very bouncy metal spring: there is some transfer of information, but the physical properties of the medium make it much harder to perceive the distal properties we are trying to gauge.

6.5 Narcissistic sensory systems

The most basic requirement for sensory representation as presented in chapters 4 and 5 is that there is a correlation between the occurrence of the representation and what's represented. In temperature perception this correlation is realized by the thermoreceptors. Thermoreceptor firing rates correlate with certain skin temperatures, and this correlation is what underlies much of the representational power of the sense modality.

This view of thermoception (and sensory systems in general) has been disputed by Akins (1996). Akins argues that the naturalistic project of explaining intentionality (or in a more limited case, sensory representation) uses strong underlying assumptions of how sensory systems work – and those assumptions are false.

The naturalistic project according to Akins aims to explain the *aboutness* (or intentionality) of representational states that is in accord with science, does not use further semantic predicates, and explains the aboutness relation *as* a natural relation such as causation or carrying information. The naturalistic project assigns a special importance to perception, as perceptual representation is the most basic case of the aboutness relation that stands in need of explanation. The theory of representation put forward in chapter 4 is an example of that naturalistic project: an attempt to reduce the aboutness relation of sensory states to the natural relation of *carrying information* and the natural phenomenon of *biological function.*

The champions of the naturalistic project, Akins claims, have in common a 'traditional' view of the senses. The traditional view of the senses has three characterizing traits:

1. Sensory signals must present a 'veridical picture' of the world through *constant correlation* between signal and stimulus.
2. Besides correlation, sensory signals must *mirror the structure* of external events.
3. A sensory modality must present the veridical picture "without exaggeration or omission" (1996).

The first characteristic of the traditional view is a strong version of the correlation requirement that I discussed in chapter 4. Akins seems to take the traditional view to hold an especially strong correlation requirement, as she writes that "In nonmetaphorical terms, this aspect of sensory veridicality is usually expressed as that of *constant correlation:* if a signal is to be informative ("tell the truth"), it must be produced when and only when a particular stimulus (or stimulus set) is present."(1996, p. 343)

The second characteristic is similar to various requirements of structural representation that I have discussed in chapter 5. The traditional view according to Akins holds that preservation of structure is a requirement for sensory representation.

The third characteristic is somewhat vague. On the surface of it there seems to be significant overlap with the first characteristic: if a sense presents a veridical picture through constant correlation, then there should be no omissions or exaggeration. But something else is meant by the third characteristic. Akins introduces it by way of metaphor, stating that sensory systems should be *servile* to the brain. I take this to mean that a sensory system should provide a 'neutral' picture on which all available information (and not more) is transferred to the brain for further processing, and that goal-directed processing in the sensory system itself does not fit in the traditional view. This contrasts with the view that Akins takes (and I do too) that senses themselves not only transfer but also process information in a task-oriented manner. The traditional view is that senses do no such thing, and merely present a neutral picture to the cognitive systems that do the further processing.

Against this traditional view Akins argues that the senses do not fit these characterizations. Using thermoception as an example, she argues that it fails (1) because firing rates of thermoreceptors in the skin do not linearly correlate with skin temperature. The same receptor response can occur at different skin temperatures, and different receptor responses can occur at the same skin temperature. Because of this equivocation, Akins claims (2) can also not be satisfied. Furthermore, the senses 'embroider' on their account of the world; receptor responses can be exaggerated in accord with the "interests and sensitivities" of body parts.(1996, p. 352) This does not fit with characteristic (3).

To replace the traditional picture of the senses, Akins proposes a view of senses as *narcissistic* systems. The way the senses react to stimuli is not the neutral veridical information transfer of the traditional picture, but rather a self-interested response to the environment. Senses are the evolved answers to specific informational needs of the systems that drive action and behavior. Motor systems usually do not need the type of lossless, undistorted information that the traditional picture presupposes, so the senses do not provide that information. The traditional picture conflates the question "what is that sense representing" with "what stimuli are the receptors responding to". Rather, we should ask "what is that sense *doing".* The answers to this question will elucidate how senses provide cognitive and motor systems with the information they need.

To Akins, the narcissistic view of the senses is a non-representationalist view of the senses. She identifies the representationalist framework with the traditional view, and based on the arguments against the traditional view, she rejects the representationalist project. For Akins, perception is not the starting point for a theory of aboutness. Aboutness is something cognitive states can possess, but perceptions usually do not. I disagree with this conclusion, but I think the disagreement is at least partly semantic in nature.

My disagreement with Akins' conclusion is that I think the narcissistic character of perceptual states does not mean they aren't representational. The theory of representation outlined in chapter 4 is more sophisticated than the traditional view Akins criticizes. On the more sophisticated theory of representation, the question of *what a state is doing* plays an important role in determining its intentional content. The content of a perceptual state is not its input conditions, nor is it the information it carries about properties in the world. The content is only that part of the carried information that explains how the system is performing its task function. Akins' idea of asking *what the system is doing* is built into the definition of representation. I think the disagreement is mostly semantic: I don't think Akins would deny that perceptual states carry information, or that carrying information can be part of the explanation of the functioning of the system – the disagreement is over whether to call it representation, and whether the task-function dependent notion of content is objective enough for 'aboutness'.

A further point about whether perceptual representation is a good starting point for a comprehensive explanation of aboutness is beyond the scope of this thesis. The traditional view had an 'easy' solution: perceptual states are about properties in the world and the contents of perceptual states truthfully refer to actual external states of affairs. The task-function based account does not offer as easy a route. Sure, there are still perceptual contents that refer to states of affairs external to the organisms (i.e., perceptual states carry information), but they aren't necessarily the neat true propositional contents that the traditional view postulated. The task-relative version of perceptual content is a more plausible theory of perceptual representation, but it doesn't offer the neat path to a naturalized epistemology that the traditional view did.

It is worth noting that Akins' characterization of the traditional view is probably mostly aimed at the modular representationalism of the 1980s and 1990s. There are many more current representationalist theories that do not endorse the premises that Akins ascribes to the traditional view, and therefore are less susceptible to this line of critique. For example, Matthen conceives of sensory systems as sorting machines that assign distal stimuli to classes which are "constructed on the basis of commonalities found or imposed by the system, not passively received". (Matthen, 2005, p. xi) Matthen's view violates the third characteristic of the traditional view: the sorting of stimuli into classes depends on the goals the system has with these categorizations, and not only on the similarities between the distal stimuli. To give another example, Nanay (2013) formulates a theory of action that appeals to perceptual states as mediating between sensory input and motor output. These perceptual states, which he calls `pragmatic representations', represent those properties of the environment that are necessary for the performance of an action and represent them in an action-relative way. To pick up an object you need a representation of its size and weight not in absolute terms, but relative to your strength, hand size, etc. The sensory system then is not necessarily in the business of providing a neutral veridical picture of the world, but rather it is solving a problem of sensorimotor control. These two examples are meant to show that the characteristics Akins ascribes to the traditional view are certainly not universally accepted in more recent theories of mental representation.

In the next section I am going to engage with problems for the correlational account of temperature representation.

6.6 Problems for correlational temperature perception

In chapter 4 I presented an account of perceptual representation that depends on perceptual states carrying correlational information about the things they represent. Akins (1996) argues that temperature perceptions do not carry the required correlational information. Gray (2013b) argues that receptor responses don't correlate to skin temperatures but do correlate to the rate of exchange of thermal energy between the skin and the environment (or vice versa). In this section I will discuss whether such arguments against correlation are knock-down arguments against a correlational-informational account of representation of the type described in chapter 4.

Static response curves
The static response of warm and cool thermoreceptors (see section 2.2) is roughly bell-shaped, meaning that for each firing rate of cool receptors or warm receptors, there are two static temperatures that could cause that firing rate. This means there is equivocation from receptor responses to temperatures. I will go through a list of responses to this problem.

1. Information from other thermoreceptor types may be used to eliminate the equivocation. At the two temperature values that have equal warm receptor

response, cool receptor response will differ. So, the temperatures associated with the same warm receptor response can be distinguished by integrating cool receptor response. Akins discusses this response to the problem of correlation and has a counterargument. Akins claim is that this way of eliminating equivocation wouldn't work, since it would require the summing of cool and warm receptor responses, and the result of that summing would depend on the ratio of warm versus cool receptors present at the location of the stimulus. If there were many warm receptors present, the sum would skew warm and if there were many cool receptors, the sum would skew 'cool'. There is great variation in this ratio of warm versus cool receptors throughout the body. To accurately sum, the summing system (i.e., some part of the CNS) would have to weigh the warm and cool responses by the ratio of occurrence of the receptors at that specific location. It is unlikely that this information about receptor ratios is available to the summing system, so it is hard to see how summing warm and cool response could lead to the elimination of this equivocation. One option Akins does not consider is a simple system where the equivocation is solved not by summing but simply by whether there is a warm receptor signal present at all. A recent study in mice suggests this may be the actual mechanism by which warm and cool are distinguished. Mice, like humans, have an ability to distinguish warm from cool and perceive warming of the forepaws at changes $\geq 1°C$. Paricio-Montesinos et al. (2020) show that the ability of mice to distinguish warm from cool is not only dependent on the presence of warm-activated polymodal C-fibers, but also on the presence of warm-silenced polymodal C-fibers. Gene-knockout mice lacking the cool-sensitive TRPM-8 receptor are unable to perceive warm, and the lack of warm-sensing ability is associated with a lack of warm-silenced polymodal C-fibers. This is evidence that mice use two populations of polymodal C-fibers (warm-activated and warm-silenced) to perceive warmth. The fact that mice can distinguish warm from cool is explained by the fact that only warm stimuli activate the warm-activated fibers and silence the warm-silenced fibers. Cool stimuli do not elicit this pattern of activation in the two C-fiber populations.

2. Another response to the correlation problem is to simply say we don't understand *how* the correlation is realized, but to insist that it *must* be. The behavioral abilities of humans (and other mammals) show that we don't usually confuse warm stimuli with cool stimuli or vice versa. Mix-ups may happen under some circumstances, but overall, we are adept at distinguishing warm from cool. Even if we grant Akins' point that a correlation is not realized at the level of the individual receptor, the behavioral evidence shows that there *is* a correlation realized at some higher neural level: how else could our behavior be reliably correlated with warm or cool stimuli? The results from Paricio-Montesinos et al. (2020) go some way to answering the question of how that correlation is realized, although it does not present the full story of how the signals from these two receptor populations are processed in the CNS. That we don't know (yet) exactly how the correlation is realized in the CNS, does not mean it isn't.
3. A third response is that Akins' requirements of correlation are too strong. In her characterization of the traditional view, the correlation requirement is formulated as: "if a signal is to be informative ("tell the truth"), it must be produced when and only when a particular stimulus (or stimulus set) is present." (1996, p. 343) By

attributing such a strong requirement to the traditional view, Akins is setting that view up to fail. Exceptionless correlations are probably non-existent in biological and neurological systems. The informational-correlational account of representation has much less stringent requirements for representation. The lower bound for representation in an informational account of representation is that the signal carries *some* information. This is a very weak requirement: even the weakest correlation carries *some* information, and thus could potentially be exploited for representational purposes. Even if we grant Akins that the correlation between stimuli temperatures and receptor response is less than exceptionless, we can still grant it representational status based on a weaker link.

4. Something else we do that Akins doesn't account for, is that in instances where we use the thermoceptive system for clearly perceptual purposes we can behaviorally compensate for some of the biases. Parents, when preparing baby formula, will often gauge the temperature with the inside of the arm rather than with the hand – doing this provides a more reliable estimate of the temperature. When you feel something to be very hot but at the same time know you have cold hands, then you may *know* your perception is unreliable and find some other way to gauge the temperature or otherwise adapt your behavior to this fact. Akins focuses on the thermoreceptors, but in processing (conscious or not) some quirks of the receptors may be ironed out.

Dynamic response curves

Besides the issue with the bell-shaped static response curve of the warm and cool receptors, there is an issue with the dynamic responses of these receptors. As explained in chapter 2, warm and cool receptors have increased responses to respectively warming and cooling.

This dynamic response can be taken as an argument against the position that thermoception represents temperatures; the dynamic responses mean weaker correlation between temperatures and receptor responses, so that the system likely fails the first requirement Akins ascribes to the traditional view. Moreover, the system seems not only to fail correlation, but to systematically distort – it consistently overreacts to warming or cooling stimuli. This fails Akins' third traditional requirement: a system is not to embroider on the information available to it. The thermoceptive system seems to do so even at the level of receptors by more extremely representing changes in temperature.

Together, these two traditional requirements form a challenge for my temperature representation model. How can thermoceptive states represent temperatures if they (a) don't reliably correlate with temperatures and (b) seem to *purposefully* deviate from accurate temperature representation in a systematic way?

The response to the first challenge is essentially given above, in the context of the static response curves of thermoreceptors. Answers (2) and (3) can be similarly applied to the dynamic response curve, but there is a limit to this argument. The less information a state carries, the less representational work it can do. The weaker the correlation between what's represented and the representation, the weaker the power of the system to provide information to the organism. Representations are states that explain the performance of a task function, but the strength of that explanation depends in part on the amount of information carried. Weak correlations simply aren't as useful as strong correlations, so there may be a point where one could argue that although a state carries *some* information about some environmental stimulus, it simply carries too little to explain the performance of the sensory system in question. With thermal receptors, this doesn't seem the case. It is clear that signals originating in warm and cool spots explain the ability of mammals to respond to thermal stimuli. Philosophical arguments about the degree of correlation between the stimuli and receptor responses cannot change that empirical fact – and this is what lends support to answer (2) above: the receptor responses clearly *are* how we perceive temperature. The task at hand isn't to philosophize how they fail, but rather to figure out on a neural level how the receptor information is processed to allow humans to thermoregulate and behaviorally respond to thermal stimuli. This is something Akins clearly sees: her call to action isn't to abandon perception science because of the failure of the classical view, but rather to focus on the question what a system is doing – which I interpret in part as the question *how does the system do what it does?*

The question of what the system is doing also informs my answer to the challenge of dynamic receptor response. In dynamic responses, thermoreceptors systematically deviate from correlating to system temperature. The problem here isn't just that this weakens correlation, but that systematic deviation seems to suggest the system is doing something else than reporting temperature.

My response to this challenge is that the dynamic response of thermoreceptor is a *feature* of the system rather than a bug. The dynamic response is an early warning system for extreme temperatures. A skin temperature that is dropping quickly signals an environmental change that needs to be compensated for. By giving an early warning, the system can adapt quicker. The fact that dynamic response is transient (it adapts) is explained by the system returning to its normal role of reporting actual temperature after the warning has been given. So, does it violate correlation? Yes, but it carries information with specific task-function goals, so it should still be viewed in the framework of temperature representation. What the system is doing is reporting *anticipated* temperatures. The answer to the question of *how* the system is doing this involves a responsiveness of the receptors not only to static temperatures, but additionally to temperature change.

The thermoceptive system is not primarily in the business of accurately reporting temperatures. It is in the business of thermoregulation, and to do so it reports temperatures in a way that suits that goal. To facilitate a timely response to anticipated high or low temperatures, the system exaggerates current inputs if they possibly point to future extreme inputs.

Receptor Distribution

Another way correlations between stimulus temperature and temperature representations fail is by the variable distribution of thermoreceptors over the skin. There is variability in two dimensions: the degree to which a particular area of skin is innervated with thermoreceptors per se, and the relative occurrence of warm vs. cool receptors. This results in differential sensitivity for warm and cold stimuli at different locations on the body. For cool stimuli in particular, some areas of the body are more sensitive than others. (Gerrett, Ouzzahra, Redortier, Voelcker, & Havenith, 2015; Ouzzahra, Havenith, & Redortier, 2012)

Akins mentions this differential distribution as a reason why the traditional view doesn't work: if temperature sensations accurately represented temperatures, then there couldn't be variability across the body, since that means the same stimulus temperature is represented differently depending on the location of the stimulus.

Again, from the 'narcissistic' perspective of thermoregulation, the variable sensitivity makes some sense. The head and torso are more sensitive, while extremities are less so. This corresponds with the importance of maintaining homeostatic temperature in the organs located in the head and torso. It also corresponds to the likeliness of temperature variability: the extremities are simply more likely to cool off sometimes (within acceptable limits), so extreme sensitivity in those areas would probably not be beneficial. This variable sensitivity does not seem very useful in the perception of the temperature of external objects: the hands are not especially sensitive compared to other body parts, even though they are packed with receptors.

It must be noted that the functions I attribute to the variable distribution are highly speculative. The variability isn't enormous and differs from person to person. It is entirely possible that some areas will turn out to be extra sensitive or less sensitive in non-obvious ways. Still, I think it makes sense to think of the increased sensitivity on the head and torso along the same lines as I argued about dynamic responses. These areas of the body are extra important to thermoregulation, and therefore exaggeration of the cool or warm response in these areas could help make the thermoregulatory system quicker at responding to relevant stimuli.

Spatial summing

In chapter 3 I flagged that the fact that some thermoreceptors exhibit spatial summing is a problem for the simple Temperature Representation Model. It is also a problem for the more sophisticated model that I have put forward in this chapter and the preceding chapters. Larger stimuli are perceived as being hotter (and not larger) even when they are in fact the same temperature. This doesn't seem to fit well with a theory that says temperature perceptions are representations of temperatures, because on that theory this would count as a systematic misrepresentation (it could even be called an illusion).

But perhaps there is an avenue of explanation: temperature perception in general has relatively poor spatial resolution. It is hard to perceive the precise boundaries of a thermal stimulus. Maybe, because of this, the importance of spatial information is downplayed by the system. The increase of receptor activity in a general region is taken by the thermoceptive system to signal a general increase in temperature in that region, rather than as a change in the spatial properties of the thermal stimulus. But this is a speculative explanation. As said in chapter 3, the Heat Exchange Model has a more straightforward explanation for this phenomenon, and I concede that that is an asset for that account.

Thermal conductivity

Also in chapter 3, I discussed the misrepresentation of temperature that occurs on the simple TRM with regards to thermally conductive stimuli. Objects that are thermally conductive feel cooler to the touch than objects that are thermally non-conductive. When they are at the same temperature, a block of steel will feel significantly colder than a block of wood. It is a challenge for my model of temperature representation.

I think we can make sense of this fact in various ways. First, if you think of thermal sensations as representations of *skin* temperatures, then it is right in representing conductive stimuli as being colder; more thermally conductive objects cool the skin more rapidly, and consequently produce lower skin temperatures at a given time.

But, as I argued earlier in this chapter, sometimes when we touch an object, temperatures are attributed not to the skin, but to the object. In such cases, there seem to be systematic misrepresentations of temperature dependent on the thermal conductivity of the object. Cases like this need a different explanation. The speculatory explanation I have to offer is that the responsiveness to thermal conductivity is an early warning system of the same type as dynamic receptor response. When a cool thermally conductive object is touched, the skin will cool rapidly, and a skin temperature that is dropping quickly signals an environmental change that needs to be compensated for. Just like with dynamic receptor response, the benefit of an early warning is timely compensation for impeding low skin temperature.

In fact, the dynamic response curve and the thermal conductivity effect are not independent phenomena: thermally conductive stimuli can warm or cool the skin more quickly, which elicits a stronger dynamic response from the thermoreceptors. The explanation for the perception of thermally conductive stimuli is the same as the explanation for dynamic response curves because dynamic response curves are why thermally conductive stimuli feel warmer or colder.

A third explanation of the conductivity effect builds on the fact that this phenomenon helps in recognizing wetness. (Filingeri, Redortier, Hodder, & Havenith, 2013) Exaggerated responses to conductive objects can aid tactile recognition of wetness, which is an important skill for land-dwelling mammals. Feeling wetness is a cross modal process which uses touch information and thermoceptive information. The exaggerated cool response to cool conductive objects is a problem for a correlational account of temperature perception, but it is at the same time a feature of this cross-modal process. By effectively amplifying the 'cool' signal, the system for recognizing wetness becomes more acute.

The common thread in my arguments on receptor distribution, dynamic responses, summing, and conductivity is this: the 'failed' correlation still manages to elicit a proper response from the thermoceptive system. The quantity in the world that the system is tracking is still temperature – even if it is temperature over time (as in dynamic responses) or temperature over an area (as in summing). The goal of the system is still (always) to maintain homeostatic temperature. So, both the inputs and outputs of the system are directly temperature-related, and the functioning of the system vitally depends on the temperature receptors responding to temperatures the way they do. To me, this clearly shows that this is a system of temperature representation, even if the correlations it uses aren't always of the same strength, and even if it may use the signals in a variety of ways.

6.7 Structural temperature representation

In chapter 5 I set out how the preservation of structure can aid representational systems. I argued that preservation of structure is not necessary for systems to be representational, but that a specific kind of preservation of structure can allow a representational system to represent structural properties such as ordered properties and quantities.

The question at hand in this section is whether the human thermoceptive system is such a system that preserves structures. A complete answer to that question would require a complete map of the representational scheme – in effect a complete description of the thermoceptive system's flow of information. Such a map is not available – so I must do my best to answer the question in a more limited sense, with the information available. My arguments will provide reasons to think there *probably* is preservation of structure, but I won't identify the structures of the representational system that preserve the structure of the stimulus properties. That task is an empirical one, and a very difficult one. An example of the empirical discovery of the preservation of structure in a perceptual system would be the retinotopic maps in vision – spatial relations between the stimuli are preserved in spatial relations between neurons in the primary visual cortex. Evidence for structural similarity on a neural level is not enough to satisfy the criteria set out in chapter 5: it must also be shown that the system *exploits* structural features of the vehicles to gain information about the relations between stimuli. One can imagine the difficulty of this task: detailed neural imaging is needed, as well as behavioral/functional evidence that the neural structure explains the functioning of the system. Direct evidence of this kind is not available for thermoception.

What we can do is reason back from behavioral abilities of humans to how the perceptual system might provide us with the relevant information that underlies these abilities. In particular, the ability to compare temperatures suggests that ordering is preserved.

It is well within the normal thermoceptive capabilities of humans to determine one object to be warmer (or cooler) than another object by touching them. This ability of course manifests itself in a limited range of stimuli (not for very hot or very cold objects) and there is a threshold below which the difference is too slight to reliably perceive.

A purely correlational system, with no preservation of structure, would not be able to tell how stimuli relate to each other in the *hotter than* relation. This ordering relation that holds between object temperature must somehow be encoded in the perceptual system for us to be able to make these comparisons. As argued in chapter 5, the system must exhibit *structural correspondence* for such comparisons to be successful.

The example of comparing temperatures not only shows that some relation is preserved in the representational system, but also that this preservation of structure can be exploited by the system itself and downstream consumer systems. The fact that I can consciously make comparative temperature *judgements* and put those into linguistic form shows that downstream systems are fed this information that is based on the structural correspondence.

According to Shea (2018), one hallmark of true structural representation is that novel intermediate stimuli get their meaning based on the position in the representational structure. Is this the case for temperature representations? The problem with answering this question is that there aren't really 'novel' temperatures. It is likely that *if* there were an in-between temperature that hasn't been encountered before by the organism or any ancestor in the selection history of the trait, then that intermediate temperature would trigger an in-between response just because of the physiology of temperature receptors. The response curves of temperature receptors are continuous gradients, not random values. So, a 'novel' intermediate temperature would elicit the intermediate response from the receptors, and consequently take the intermediate place in the representational structure.

To what extent structure is preserved in temperature perception and how that preservation of structure is realized in the central nervous system is an interesting avenue for further thermoception research.

6.8 The spinothalamocortical pathway

In chapter 2 I mentioned a recent development in the neuroscience of thermoception: the discovery that in rats, thermoregulation is largely driven by signals mediated by the lateral parabrachial nucleus (the LPB pathway) and not by the spinothalamocortical pathway, which is responsible for our ability to discriminate warm and cold temperatures. Even *behavioral* thermoregulation in rats seems to be mediated by the LPB pathway and not by the thalamic pathway, as evidenced by experiments with thalamic-lesioned rats placed on an array of plates that ranged from uncomfortably cold via thermally neutral to uncomfortably warm. The thalamic-lesioned rats still succeeded in finding the plates that had a comfortable temperature and avoided plates with uncomfortable temperatures. This shows that heat- and cold avoidance remained intact even if the pathway responsible for the discrimination of temperatures of external objects was functionally ablated. (Yahiro et al., 2017)

If this result carries over to the human case, this could mean that conscious perceptual states are not (always) what leads to thermoregulatory behavior. On a commonsense view of conscious thermoregulatory behavior, one might think that it goes something like this: first, a person has a conscious thermal experience (e.g.: it's cold outside), and then that mental state leads the person to engage in thermoregulatory behavior (such as going indoors).

In this thesis I have been assuming that conscious perceptions of temperatures are not *distinct* from thermoregulation, but rather a part of the human thermoregulatory system. The two-pathways finding could be taken to suggest that the ability to consciously perceive and distinguish temperatures is quite distinct from the mechanisms that drive thermoregulatory behavior. However, I have some doubts that the results from this experiment translate neatly to the human situation.

The thermal behavior that the rats exhibit in the experiment is relatively simple behavior – both the stimuli and the resulting behavior are not very complex. Humans have a much more diverse set of thermoregulatory behaviors they can exhibit than rats, and they usually also have more freedom of choice than the rats in this experiment. The rats in this experiment are responding to stimuli that would lead to hyperthermia (or hypothermia) if sustained too long. It would be interesting to see if the same results would be obtained with the plates at temperatures closer together – it is possible that in a setup that needs more fine-grained discrimination, the thalamic pathway plays a role.

More complex human thermoregulatory behavior likely *is* mediated by more complex central nervous processes. Taking off a jacket is a relatively simple thermoregulatory action, but if you are in a situation where doing so would be inappropriate, you may opt for another thermoregulatory behavior instead (say, opening a window). Or perhaps it is a hot day, and you decide to pick up a bag of ice at the store on your way home so that you can enjoy a cold drink later: this is a complex behavior with no immediate thermoceptive results; it seems unlikely this is mediated by the LPB pathway exclusively. Perhaps in more complex behaviors, or situations that require deliberation, conscious perceptual information mediated by the spinothalamocortical pathway plays a larger role than in the experimental setup used by Yahiro et al.

If it does turn out that human thermoregulatory behavior uses a strongly distinct pathway from temperature perception, this would mean that the object-temperature attributing mode of thermoception is perhaps more distinct from thermoregulation than I have described it at the beginning of this chapter. Still, it uses the same set of receptors and responds to the same physical quantity in the environment – the modes remain highly interdependent.

6.9 Temperature illusions

In the following section I will discuss a few temperature illusions – situations in which the thermoceptive system functions differently from our expectations and does so in a systematic way.

Three bowl illusion

Perhaps the most well-known temperature illusion is the three bowl experiment. It is discussed in Locke's Essay Concerning Human Understanding. (Locke, 1690, 2.8.21) The experiment is often wrongly credited to E.H. Weber. The experiment goes like this: one hand is placed for several minutes in a bowl of warm water of 40°C, and the other in a bowl of cold water of 10°C. Then both hands are placed in a bowl of tepid water at 27°C. The resulting experience is that the tepid water feels distinctly warm to the hand that was submerged in cool water, and distinctly cool to the hand that was submerged in warm water. (Tritsch, 1990) This result is surprising because of our general ability to make relatively reliable replicable temperature judgements in the range around 27°C.

The suggestion gleaned from the experiment is that temperature sensation is strongly dependent on adaptation temperature, which poses a problem for my theory of temperature perception in a way that is similar to the problem of receptor distribution. If temperature perception depends strongly on the prior temperature of the skin, then is it a reliable representation of temperatures of objects in the environment?

Tritsch (1990) showed that the occurrence of the illusion depends strongly on the specifics of the experimental setup. In performing the experiment with various temperatures for the tepid water, it was shown that the range in which the illusion occurs is limited. Similar experiments with solid temperature probes that were gripped with the hand (rather than bowls of water) yielded an even narrower range in which the illusion occurs. If solid probes are only touched with the tips of the fingers the illusion is even less apparent. These results suggest that the three bowl illusion is a special case, not a general failing of the thermoceptive system. In fact, there is some evidence that shows an effect *opposite* to the normal three bowl illusion. Egeth (1970) had subjects place their hands on two plates of different temperature, e.g., the left hand at 35°C and the right at 32°C. After an adaptation period, the hands are placed of two plates of the same temperature, e.g., 40°C. If the 'classic' three bowl experiment generalizes, you would (correctly) expect the right hand to feel the 40°C plate as warmer than the left hand does. This is called *contrast*: the enhancement of subjective intensity when the stimulus follows a weaker stimulus.[5] But at some temperature pairings and adaptation times, an opposite result was measured: the equivalent of the *left*-hand plate feeling warmer in the setup described above. This is the opposite of contrast: assimilation.

My explanation for the three bowl illusion is that it is a consequence of the dynamic receptor response: because the temperature increase in the pre-cooled hand is larger, the dynamic response of the warm receptors is stronger – and vice versa for the pre-warmed hand. As discussed above, the dynamic response of thermoreceptors is a *feature* of the thermoceptive system, not a bug. The fact that this feature makes for misrepresentation within a limited range of cases (only for specific stimulus types and temperatures) does not pose much of a problem for the representational theory of temperature perception. Furthermore, the difference in the strength of the illusion between the wet stimuli and dry stimuli showed by Tritsch (1990) suggests that the dynamic response involved in recognizing materials (based on thermal conductivity) is a feature of the system that may be responsible for the illusory experience in the three-bowl scenario. What the three bowl experiment shows is how the multiple functions of the system (temperature identification, temperature change response, and material identification) can sometimes lead to conflicting goals within the system, that can be brought out in specific circumstances.

Silver Thaler Illusion

The Silver Thaler Illusion, also known as Weber's effect, is the effect that a warm or cool object placed (e.g., a silver thaler) on the skin will feel heavier than a thermally neutral object of the same weight and size. (Stevens, 1979, 1982; Weber, 1978)

While this is an illusion induced by a thermal stimulus, it is not a temperature illusion. The property that we are mistaken about is the weight of the object, not the temperature – as such it is a tactile illusion induced by a thermal cue. I do not see the silver thaler illusion as a problem for theories of thermoception. The effect is likely due to the cold sensitivity of certain mechanoreceptors. (Cahusac & Noyce, 2007) The silver thaler illusion is closely related to the phenomenon that tactile spatial acuity is greater for cold stimuli than for neutral stimuli. (Stevens, 1982)

---

[5] Contrast is also the diminution of subjective intensity when the stimulus follows a relatively strong stimulus.

Paradoxical cold

Paradoxical cold is the phenomenon that some noxiously hot stimuli can feel cold under certain circumstances. (Campero et al., 2001; D R Kenshalo & Duclaux, 1977; Long, 1977) It is a temperature illusion because there is a surprising systematic misrepresentation of temperature that occurs under specific conditions. The phenomenon is due to some cold fibers being activated by stimuli in the noxious hot range. (Campero et al., 2001) The effect is dependent on body temperature, or vasoconstriction more generally (Long, 1977). This phenomenon is relatively rare, and I don't have an explanation in terms of system function. I accept this illusion as a quirk of the thermoceptive system that awaits further explanation both on a molecular/receptor level and on a system-function level.

This illusion is not well-explained as a *narcissistic* trait of the sensory system either, as the deviation from correlation seems to have no benefit. The Heat Exchange Model does not provide a ready explanation for this phenomenon either: feeling a noxiously hot stimulus as cool not only gets the temperature wrong but also gets the direction and amount of heat transfer wrong.

Chemical sensitivity of thermoreceptors

Because some thermoreceptors are sensitive to chemical stimuli, some compounds can elicit thermal sensations in the absence of the relevant temperature stimulus. For example, chili peppers are perceived as being painfully hot due to the capsaicin they contain which activates TRPV-1. Menthol is perceived as cool due to the activation of the cold-sensing ion channel TRPM-8. This phenomenon isn't usually described as an illusion, maybe because we are so familiar with it. It is, however, a surprising systematic misrepresentation of temperature. I do not have a functional explanation of this illusion – it is simply a fact of the molecular structure of thermoreceptors that they can be activated by certain chemical stimuli. This fact is sometimes evolutionarily exploited by plants as a defense mechanism: the capsaicin in the fruits of the plant genus *Capsicum* is likely a defense mechanism against grazing by mammals. Birds are not sensitive to capsaicin and are the primary dispersers of seed for the plant. (Tewksbury & Nabhan, 2001)

7 Towards a representational theory of thermal pain

Until this point in the thesis, I have focused on innocuous temperature sensations, only briefly touching on painful temperature experiences in the discussion of temperature receptors. Painful sensations of noxious heat and noxious cold are an integral part of the thermoceptive system and play an important role in thermoregulation.

This chapter presents an evaluativist account of thermal pain. Evaluativism about pain is the idea that pain represents a bodily harm or disturbance and represents that disturbance as bad for you. Evaluativism is a representational theory of pain, with an added feature that explains pain's unique phenomenology and its directly motivating force. In this chapter I give an evaluativist explanation of thermal pain and show how evaluativism relates to the task-function account of representation for non-painful thermoception.

Section 7.1 offers a brief overview of how nociceptors in the skin contribute to the experience of painful heat and cold. Section 7.2 introduces the evaluativist theory of pain. Section 7.3 is dedicated to showing how thermal pain is especially well described by an evaluativist theory of pain. Section 7.4 generalizes some points from the previous section to the case of non-painful thermoception. Section 7.5 discusses an alternative theory of thermal pain: the intensive stimulus theory. I argue that the homeostatic evaluativist theory I present is superior. Section 7.6 considers some implications of the theory presented.

7.1 Nociceptors and noxious heat experience

As mentioned in chapter 2, there are four main classes of thermoreceptors in the skin: innocuous warm, innocuous cool, noxious hot, and noxious cold. These last two are nociceptors.

Cold nociceptors respond to skin temperatures below 20°C, and the intensity of cold pain increases linearly until a temperature of 0°C. The receptors for noxious cold are inactive at normal skin temperatures, so they don't contribute to the sensation of innocuous coolness. The receptors for innocuous coolness are mostly inactive at the noxious low temperatures, so these don't substantially contribute to the sensation of cold pain. In gene knock-out mice, knocking out the main receptor protein (TRPM8) for innocuous cold has much less effect on noxious cold sensation than it does on innocuous cold sensation. That is to say that receptor molecules for innocuous coolness contribute little to the experience of noxious cold. (Schepers & Ringkamp, 2010)

Hot nociceptors come in two main types: fast conducting A∂ fibers and slow conducting C fibers. The difference in conductance speed is responsible for the distinct sensations of a fast 'sharp' pain (onset after ca. 0.4s) and a slower dull pain (onset after 1-2s) that are felt in heat nociception. A-fibers come in two types as well. Type I fibers respond slowly (peak after 5s) to prolonged stimuli of >53°C, and this response increases with longer stimulation. Type II fibers respond immediately (<1s) to stimuli of >47°C, and the response adapts to longer stimulation. The C-fibers involved in heat nociception have various thresholds (37-49°C). The heat pain threshold lies somewhere around 45°C, so the pain signals from C-fibers that are responsive to lower temperatures are likely subject to temporal summing.

For innocuous warmth and noxious heat, we can distinguish dedicated afferents, but there is a subclass of afferents for innocuous warmth that also contribute to noxious heat sensation. The peak discharge rate of that subclass of afferents lies in the noxious range. One of the receptor proteins responsible for innocuous warmth sensation (TRPV3) also plays a role in noxious heat sensation. This means it's likely that receptors for innocuous warm sensation also contribute to the experience of painful heat. Some receptors for innocuous coolness can respond to stimulation with noxious heat, which can result in a paradoxical cold sensation. (Schepers & Ringkamp, 2010)

For cold pain, colder temperatures are more painful than higher ones. For heat pain, higher temperatures yield more intense pain. Temperatures that are on the edge of painfully cold or painfully warm may elicit mild pain *and* 'normal' thermal sensations. The nature of thermal stimuli to the skin makes it so that usually when an area is heated or chilled to noxious temperatures, there is a surrounding area of skin that has a temperature in the innocuous range. Perhaps this contributes to our ability to distinguish heat pain from cold pain (although this ability is not infallible).

7.2 Evaluativist accounts of pain

Philosophical theories of pain can be broadly divided into two categories: those that see pain as representational, and those that do not. If pain is representational, that means that it is (theoretically) possible for pain to misrepresent. Representational theories of pain tend to liken it to sense modalities (such as vision), arguing that nociception is a sense that perceives a state of the body, specifically damage or harm done to the body. Representational theories tend to put more weight on this aspect of pain as a signaling system for bodily harm – pain indicates damage or impending damage. (Armstrong, 1962; Shoemaker, 2000; Tye, 2005)

Non-representational theories of pain put an emphasis on the experiential character of pain: there is a sense in which a person cannot be mistaken about being in pain – when you feel pain, then you have pain. This idea that pain experience is infallible is taken as an argument that it isn't representational since representation implies the possibility of misrepresentation.

Another aspect that non-representational theories allude to is that pain is directly motivating; that something will hurt is a reason not to do it. We intrinsically want to avoid not just the cause of pain, but also pain itself. On the representational view, perhaps, there needs to be a further reason to not want to represent bodily damage. If the representationalist is right, and a pain experience is a useful piece of perception, why would we want to *not feel it?* Is it rational to take painkillers on a representational theory of pain? (Aydede, 2009; Block, 2005; McGinn, 1982)

Evaluativist accounts of pain are developed with the aim of giving a representational theory of pain that accounts for the directly motivating aspect of pain and for the rationality of pain avoidance and pain relief.

According to evaluativist accounts, pain represents some bodily condition, and represents this condition as *bad for you.* This is the *evaluation* part of the evaluativist account. The idea that pain represents a bodily condition is meant to cover the first aspect of pain mentioned above: its role as a signaling system. In this way, the evaluativist account incorporates the picture of pain as a sense modality and as part of the somatosensory system – pain reports on the state of the body. (Bain, 2017)

Different versions of an evaluativist account can offer different answers to the question of what the nociceptive system represents. For example, an evaluativist can hold that pain states represent tissue damage, or bodily harm, or impending damage.

The *evaluation* aspect of the evaluativist account is contained in the idea that pain not only represents some bodily condition, but also represents it as *bad for you*. This is supposed to explain the strong affective aspect of pain – unlike 'normal' perception experiences, pain is always linked with a specific type of negative affect. The phenomenology of pain sets it apart from our regular sense experiences. (Bain, 2017)

The evaluation aspect also provides a basis for the directly motivating character of pain. Because of its negative affect, pain *is* a reason (not) to do something. Avoiding pain or alleviating pain is a reason for an actor to do something, and that reason is not always mediated by a further desire. The reason to perform a certain action (or not perform it) may be simply to avoid pain. That pain is inherently unpleasant also explains why we would want to avoid pain – not just the bodily harm it represents, but the experience of pain itself.

So, for the evaluativist, representing the pain as *bad for you* explains why we avoid pain itself, and not just the causes of it (e.g., by taking painkillers) and it explains how pain can provide a direct motivation for action. (Bain, 2019)

On an evaluativist account, pain has representational content: pain represents a bodily condition. For there to be representational content, there must be a possibility of misrepresentation. This goes against the intuition that pain is infallible – that you can't be mistaken about whether you are in pain.

That means they must give a counterargument to the infallibility of pain experience. The usual counterargument given by the evaluativist is that *experience* is always infallible in a certain sense. When you have a visual illusion, although the content is false, you still have a claim to having had that experience. The same is true for pain: a false pain episode has the experiential character of a true pain episode, but you are mistaken. The perceived infallibility of pain attribution stems from the fact that we attribute pain based on experience, while other sensory experiences are usually success terms. If this difference in attribution is a merely verbal difference, and not a metaphysical one, then pain is really no different from other senses. Whether one finds this counterargument convincing probably depends on how strongly one buys into the intuitive picture of pain as infallible. I am not going to make the case for this argument in this thesis – I will just note that it's a common part of representationalist accounts.

To summarize, the evaluativist theory of pain reconciles the two aspects of pain by giving a representational theory of pain and including a component that *evaluates* the represented state *as being bad*. The intrinsically motivating aspect of pain is incorporated in a representational theory. The evaluativist must also explain how pain *is* fallible.

Another representationalist (but not evaluativist) account of pain is *imperativism*. (R. J. Hall, 2008; Klein, 2007, 2015; Martínez, 2011) According to imperativism, pain states represent an imperative to act a certain way, for example to protect the body part in pain. (Klein, 2007, 2015) Hall and Martínez develop a mixed theory, where pain states have both indicative content (along the lines of: "there is bodily harm here") and imperative content, while Klein develops a view in which pain has no indicative content. Imperativist and evaluativist theories are related in that both are representational theories that posit additional content (beyond indicative) to explain pain's peculiarities. Where evaluativism needs to explain how a bodily condition is represented as 'bad for you', an imperativist account must explain what exactly pain implores one to do. The imperativist theories mentioned above give slightly different answers to this question, and it is beyond the scope of this chapter to go into these intricacies. The next section of this chapter will present an evaluativist view of thermal pain which I think is more elegant than an imperativist view would be, and I will return to that point at the end of the following section.

7.3 An evaluativist view of thermal pain

If evaluativism about pain in general is true, it should also hold for thermal pain. In this chapter I am not making the argument that evaluativism in general is true. I will argue that a specific kind of evaluativism is a particularly apt analysis of thermal pain.

In the above paragraph on evaluativism I was purposely vague about the contents evaluativists ascribe to pain states, so that I could give a more general characterization of the view. It is common to evaluativist accounts that pain states represent some bodily condition, but what that condition is has remained unspecified so far in this chapter. It is a challenge for evaluativist to point what property is represented by pain. Cutter and Tye, for example, identify pain's representational target with bodily harm, which in turn is understood as impeding the proper functioning of the body, which in turn is given an evolutionary explanation. (Cutter & Tye, 2011) Other evaluativists choose to leave 'bodily harm' as an unexplained term, that perhaps later can be specified. (Bain, 2017)

A very interesting alternative, being developed by my colleague Magdalini Koukou, is the idea that pain states represent disturbances of homeostasis. (Koukou, n.d.) When you are in pain, there is a threat to the homeostatic integrity of your body. The notion of homeostatic disturbance is much more general than just tissue damage (which is a kind of homeostatic disturbance) and includes situations where tissue damage could *potentially* occur if the disturbance persists. I will leave the task of arguing for this flavor of evaluativism *per se* to Koukou and focus my efforts on applying it to the case of thermal pain.

Besides representing the occurrence of some bodily condition (such as a disturbance of homeostasis), pain states can have further content. For example, many pains are localized, so the content of the state may be that there is a disturbance of homeostasis at location X.

Beyond the condition attribution and localization, evaluativists claim that pain states come with an evaluative component. It is a question for evaluativists how that component stands to the simply attributive part of pain. One possible answer is that it is a higher-level property that is represented: a pain state represents a property A at location X and represents A as being F.[1] This answer, which many evaluativists seem to endorse, makes it a sort of double-representational theory. Not only does pain represent a bodily state, but it also represents that bodily state as being 'bad', or some other negative evaluation.

But this representational doubling is not necessary. The advantage of a representational theory is that it allows for misrepresentation: when you think you see a zebra while really it is a horse under a striped blanket, we can still say you represented a zebra. Saying pain is a representational state makes it possible to talk about pain as a signaling system that is prone to error. But the evaluativist component of pain does not need to have this possibility of error built in if the bodily state that the system signals is a 'negative' thing itself. Let me expand on that a bit.

A disturbance of homeostasis *is* a bodily state that necessitates a change for the system to be restored to homeostasis. The need for 'action' (anywhere from the molecular to the personal level) is contained in the concept of a non-homeostatic state in a system that is maintaining homeostasis. In other words, a disturbance of homeostasis is *always* 'bad for you'. Therefore, if a pain state correctly represents a disturbance of homeostasis, then it can't be the case that this isn't bad for you in some sense. It is true that in the long run disturbances of homeostasis can have benefits: we intentionally damage bodily tissues during surgery to cure or prevent worse conditions. But that doesn't make it incorrect that the tissue damage incurred during surgery is bad for you.

My proposal for a theory of thermal pain is that it represents an immediate threat to homeostasis, and this is in itself a directly motivating state. There is no need to *further* represent that threat to homeostasis as a threat to the organism. Thermal pain further has spatial content, and sometimes (but not always) we can tell whether the stimulus was hot or cold.

This evaluativist account of thermal pain makes sense from the viewpoint of thermoception as a system aimed at maintaining thermal homeostasis. Thermal homeostasis has multiple elements: the body must maintain a temperature within a certain range for metabolic processes to be successful, but there are also more local concerns: a very low or very high skin temperature can cause freezing or burning of the skin. The nociceptors in the skin respond to the temperatures at which such local damages can occur. The directly motivating nature of pain experiences allows animals to quickly respond (behaviorally) to these noxious thermal conditions: pain can drive responses such as a withdrawal reflex.

---

[1] Here I mean that property A has the property of F, not that A and F are identical.

The homeostatic evaluativist theory provides something like what imperativism offers, but in a more elegant way: the motivating nature of pain is explained by the appeal to its function in homeostatic systems. On the homeostatic evaluativist theory, the pain state carries content about there being a disturbance of homeostasis at a certain location, and homeostatic systems then respond to restore homeostasis. Imperativist theories posit that some signals have imperative content while other (perceptual) signals do not. So, the imperativist posits two *types* of content: one indicative (for senses) and one imperative (for pain). My theory, on the other hand, needs no such distinction. Pain states, like sensory states, have as their contents the information which helps the system perform its task function. I think of that content neither as purely indicative (since the system isn't just in the business of creating perfect representations of the state of the body) nor as purely imperative (since the system uses information *about the state of the body* in its performance of the task function).

7.4 An evaluativist account of innocuous thermoception

There is a strong parallel between the evaluativist picture of thermal pain presented in the last section, and the task-function account of innocuous temperature perception I have presented in the preceding chapters. On these accounts, both thermal pain and innocuous temperature perception function as representational systems aimed at maintaining homeostatic temperature control. On the side of innocuous temperature perception, this view deviates from a classical representational view in that it places more emphasis on what the representations are *for*. It draws a perceptual system closer to what we think of as the somatosensory system. On the side of thermal pain, an evaluativist account draws the pain system closer to a perceptual system, by maintaining that pain is representational.

This parallel between an evaluativist theory of thermal pain and a task-function theory of innocuous temperature perception suggests that there could be such a thing as an evaluativist theory of innocuous temperature perception. Many non-painful thermoceptive experiences play an evaluative role: low or high innocuous temperatures are experienced as uncomfortable, albeit not painful. Thermal discomfort is directly motivating in much the same way thermal pain is. Behavioral thermoregulation needs no further reason besides the body or a part of the body being in an uncomfortable temperature range. These two elements are recognizable from thermal pain: there is a distinct negative phenomenology (thermal discomfort), and the states are directly motivating.

The phenomenology of innocuous thermal discomfort is not the same as that of pain, and temperature-responsive nociceptors are distinct from the thermoreceptors for innocuous stimuli. I don't mean to say that innocuous thermoregulation is a type of thermal pain, or that thermal pain isn't any different from innocuous thermoception. There is a real distinction between thermal pain and innocuous thermoception: thermal pain utilizes different neural pathways both peripherally and in the CNS, and thermal pain has a phenomenology clearly different from innocuous thermoception. But the evaluative character of pain is something that is also present in non-painful thermal perception.

Temperature perceptions are never just *passive* pieces of data: they relay information that provides input for thermoregulation. In a way, each temperature perception carries with it the evaluation of whether it is something that warrants thermoregulation. That evaluation takes the form of positive or negative feedback on thermoeffectors.

Other senses do not have this feature to the same extent. Take vision: overexposure to light hurts your eyes, which, according to the evaluativist, means that there is acute or impending harm to the eyes which is evaluated as being bad. But non-painful vision is not so closely connected to the maintenance of homeostasis. Sure, vision allows animals to find food or recognize threats, which contributes to homeostasis, but it does so in a more roundabout way. Not every individual visual perception contributes to a direct homeostatic mechanism, like thermoception does.

7.5 The intensive stimulus theory

The evaluativist theory comes from a school of philosophy that likens pain to a sense modality. Historically, this view was motivated by the idea that pain was the product of a distinct system of nerve fibers. In opposition to this idea stood the *intensive stimuli theory*, which held that pain was the intensive stimulation of the nerve fibers of other senses. As neuroscience advanced, and dedicated nociceptors were identified, the intensive stimuli theory lost popularity. But a new version of it has been defended by Richard Gray, the proponent of the Heat Exchange View discussed in chapter 3. (Gray, 2014)

Gray argues that pain is not a sense modality as it fails to be singled out by the usual criteria for the individuation of sense modalities. Moreover, pain plays a different role in perception than sense modalities do. According to Gray, pain indicates when stimuli in a certain modality become too intensive. For example, too much light hurts the eye and very loud sounds hurt the ears. The pain from burns or freezing are caused by intense stimuli in the thermoceptive modality. The intensive stimuli theory distinguishes itself from theories that see pain as a sense modality: on the intensive stimuli theory, pain is a warning system that's *common* to various sense modalities which signals that a stimulus in that modality is too intense.

There is much that is right about the intensive stimuli theory. Intensive sensory stimuli can certainly cause pain. And in the case of thermal pain, it is true that the high and low extremes of the perceptual range are what elicits thermal pain. But as a general theory of pain, the intensive stimulus theory is lacking.

Modern representationalist theories of pain such as Tye's or Bain's don't literally identify pain with a sense modality, they merely *liken* it to one. Not every representational system is a sense modality, so the argument that pain isn't picked out as a sense modality by the usual criteria of individuation doesn't affect the validity of a representational theory of pain *per se*.

What the intensive stimulus theory fails to explain is why pain would be uniquely associated with harm to perceptual systems, and not to other harms. Headaches, stomach cramps or menstrual pains are not very well explained as intensive stimuli in other sense modalities, because they are interoceptive pains that don't neatly correspond to any stereotypical sense modalities. Of course, if one has a non-sparse view of sense modalities, we may consider various types of interoception to be sense modalities. But, for example, in headaches, it is hard to say what the 'normal' non-painful perception is of that sense modality.

A second issue for the intensive stimulus theory is that the neural pathways for pain are often quite distinct from the pathways for innocuous perception in the same modality. Gray argues that pain is not a separate sense modality, but rather a "functionally integrated feature of all perceptual systems". (Gray, 2014) He invokes thermal perception as an example: hot nociceptors become active at temperatures where receptors for innocuous warmth are mostly inactive, and cold nociceptors are active when innocuous cool receptors are not. Gray says that this coordinated functioning of those four receptor types shows that thermal pain is an integrated part of the temperature perception system, a part which indicates when excessive amounts of thermal energy are being exchanged.

What this argument overlooks, is that thermal nociceptors and thermoreceptors in the innocuous range are still distinct physiological structures and are wired up to the CNS in different ways. The processing of noxious thermal stimuli is very different from the processing of innocuous thermal stimuli. It can still be true that thermal pain is a warning system for intensive thermal stimuli, but it is not necessarily a warning system that is *part* of the system for innocuous thermoception. This point can be made clearer with a little thought experiment.

Imagine that humans have receptors in the retina that respond to high levels of ultraviolet (UV) radiation. The activation of these receptors causes a painful sensation in the eyes, much like looking into a bright light source does. In this scenario, human vision is not otherwise responsive to UV radiation. We can't 'see' UV, we can only feel it as pain when there are noxious levels of UV present. Would the intensive stimuli theory see UV pain as a response to an intense visual stimulus, even if it responds to a stimulus the *normal* visual system does not respond to? UV pain can function as a warning system of damage to the visual system, since UV radiation damages the retina (and other parts of the eye), but it seems wrong to say that it is a warning that visual stimuli are too intensive.

A third argument against the intensive stimuli theory is that it doesn't provide a solution or an explanation of the debate between pain as a representational sense modality, and the supposed infallibility of pain. The intensive stimuli theory claims that pain signals intensive stimuli, which probably makes it a representational theory of pain (albeit one that denies pain is a distinct sense modality). The problems with representational theories of pain were the supposed infallibility of pain, and the directly motivating character of pain. The intensive stimuli theory does little to explain either of these phenomena, so if it is to succeed as a representational theory of pain, it needs at least two additions: an argument that pain is fallible, and a reason why pain is directly motivating.

The evaluativist framework could provide these two additions. The resulting theory would be an evaluativist intensive stimulus theory: the idea that pain represents stimuli as being intensive, and that it represents this intensity as being *bad for you*. This combination would explain the directly motivating aspect of pain, but it would do so in a more roundabout way than the homeostasis version of evaluativism: a threat to homeostasis is inherently something worth avoiding, but stimuli in themselves are not. The line between non-intensive stimuli and intensive stimuli must be drawn somewhere, and an obvious place to draw it would be at the point where stimuli become potentially harmful. This harmfulness of intensive stimuli could then be cashed out in terms of disturbances to homeostasis.

In short, I think the most plausible version of the intensive stimulus theory is just a special case of the homeostatic evaluativist theory. But it is a limited version of the homeostatic evaluativist theory: one that only applies to certain types of pain caused by intensive stimuli in sense modalities. The most plausible version of the intensive stimulus theory still has trouble dealing with kinds of pain that are not obviously linked to any specific sense modality.

7.6 Conclusion and implications

I conclude that an evaluativist approach is particularly suited to provide an explanation of thermal pain, especially when pain is understood as representing disturbances to homeostasis. This theory of pain also puts the representational theory of thermoception presented earlier in this thesis in a new light: the close connection of the thermoceptive system to the thermoregulatory system allows us to analyze non-painful thermal experience through an evaluativist lens as well. The general idea that perceptual states can be evaluative has wide-reaching implications. The idea presented in this chapter is that in some cases, perceptual states can be evaluative without requiring further representation of that state as bad (or good). This makes it the case that perceptual states can provide direct motivation for behavior. This stands in opposition to a traditional theory of intentional action that holds that reasons for action are pairs of a desire and a belief, which together constitute the cause of that action. Pain states are not desires, nor are they beliefs. They provide the input information for action, and in that sense, they are belief-like. But as argued above they can also provide a direct reason for acting on that information, and in that sense, they are desire-like. Evaluativism as I have described it above sees pain states as both representational and conative, which explains how pain states can be what Nanay (2012) calls action-oriented perceptions, or pragmatic representations (Nanay, 2013): perceptual states that are the direct precursors of actions. If the evaluativist framework is indeed expandable to (some) non-painful sensory states, then that can flesh out this notion of how sensory states can guide action: by the conative component being contained within the perceptual state itself.

8 Vividness in temperature perception

Vividness is a property of mental states such as perceptions, mental images, and memories. The vividness of a mental state can be thought to be reducible to the intentional content of the mental state, or it can be a non-intentional feature of the mental state. In this chapter I argue that the application of vividness to temperature perception poses a special problem for the view that vividness depends on the intentional content of the mental state. In temperature perception it is difficult to distinguish between the intensity of the stimulus and the property represented in the perceptual state. This makes it difficult to point out which part of the intentional content of a temperature perception or mental image could determine the vividness of that state. This provides support for the idea that vividness is not reducible to intentional content.

In section 8.1 and 8.2 I will introduce the concept of vividness as it is used in philosophy and psychology. I distinguish between two claims related to vividness: the differential claim, which says that the degree of vividness can help distinguish types of mental states, and the quality claim, which only says that mental states can carry a degree of vividness. Sections 8.3. and 8.4 are about two views of what vividness is. According to the intentional content view, vividness depends on the intentional contents of the mental states. The non-intentional view denies this. Section 8.5 argues that the intentional content view is problematic when applied to temperature perception. Section 8.6 considers an alternative view of vividness that says vivid experiences have more determinate content. I conclude in section 8.7 that the intentional content view of vividness is implausible for temperature perception and imagery and discuss the philosophical consequences this may have.

8.1 Perception and mental imagery

In both philosophy and psychology, the concept of vividness has been applied to perception, mental imagery, and memory; sometimes to distinguish these categories and sometimes simply as a quality that perceptions or mental images may have. In this literature the notion of vividness is predominantly applied to visual perception and visual mental imagery, even though it can be applied to sounds, feelings, smells and other sensations or mental images.

In this chapter I discuss two notions of vividness and test the applicability of those notions to the sensory modality of temperature perception. I distinguish two ways of understanding vividness: either it is fully determined by the perceptual content of the mental state, or it is a feature mental states can have that is not fully determined by the intentional content of the state. I argue that vividness is difficult to apply to perceptions of temperature if we think of it as fully depending on the content of mental states.

I claim that in temperature perception the intensity of the stimulus is bound to the represented property of the stimulus, and that this makes it hard to identify what content of temperature perceptions could count towards vividness. This has implications for the view of vividness as depending on perceptual content, while the view of vividness as a non-intentional feature of mental states is untouched.

In the previous chapters I have argued that temperature perception can be counted as a sense modality which allows us to perceive our own body temperature, temperature of body parts, and the temperature of external objects. In this chapter I also talk about the mental imagery of temperature.[1] It is intuitively plausible that we can *imagine* certain temperature experiences, much like we can imagine a visual scene or a sound. If I ask you to imagine taking a bath that is just the right temperature, and then ask you to imagine taking a bath in ice water, it's likely that the scenes you've imagined do not differ very much visually but do differ with regards to the imagined temperature experience. While it may perhaps be difficult to have a *purely* thermoceptive mental image, it seems clear that mental imagery can have thermoceptive content. Along the same lines, I take it to be uncontroversial that we can have episodic memories that involve temperature percepts.

8.2 What is vividness?

The concept of vividness takes a place in philosophical literature as distinguishing perception, imagery, and memory. David Hume famously claimed that the difference between ideas and impressions lies in the fact that the latter have greater force or liveliness than the former. (Hume, 1739/2010, §1.1.1) In contemporary philosophy Hume's idea resonates in the claim that a difference in vividness characterizes the different phenomenology of perceiving versus imagining. (Scarry, 2006, Chapter 1; Thompson, 2014) I call the claim that there is an essential difference in vividness between various types of mental states the *differential claim*.

> Differential claim: there is an essential difference in vividness between perception and mental imagery, and between perception and episodic memory.

Most contemporary philosophers would deny that vividness provides a meaningful way of distinguishing between perception and mental imagery but may still employ the concept as a quality that imaginative episodes or perceptions can have. On this view, mental imagery is not necessarily less vivid than perception, but things can be said about the vividness of individual episodes. Basically, every philosopher or psychologist who applies the notion of vividness to perception or mental imagery must subscribe to what I call the *quality claim*.

[1] I use the term 'mental imagery' in this chapter, where some of the literature I cite uses 'imagination'. My examples, as well as the psychological measurement tools I refer to, focus on prompted imagination such as: "Imagine you are on a tropical beach". Mental imagery is a much wider category of psychological events than prompted imagination. I believe my argument holds for this broader class of mental states that can be called mental imagery, and not just for prompted imagination.

> <u>Quality claim</u>**:** mental states can have a certain degree of vividness, and the relative vividness of token mental states can be compared.

Note that anyone who endorses the differential claim must endorse the quality claim, but not the other way around.

One contemporary theory of mental imagery defines it as a quasi-perceptual process that is similar to perception in many regards except that it occurs in absence of the corresponding perceptual stimulus. (Nanay, 2017) Both the quality claim and the differential claim are in principle compatible with such a view, but since the essential difference between perception and mental imagery is given by another factor (the absence of a certain type of stimulus), this theory of mental imagery does not constitute a reason to endorse the differential claim. I endorse this definition of mental imagery, but I think my argument is likely compatible with other views of mental imagery.

This chapter does not aim to argue for or against the differential claim. Rather, I am trying to argue that the quality claim is problematic, depending on how you conceive of vividness.

In psychology the term *vividness* has also found widespread use in characterizing the phenomenology of mental imagery. Specifically, the imaginative ability of individuals is measured through questionnaires in which the subjects self-report the vividness of the mental image resulting from a given prompt. The questionnaires maybe be modality-specific or general to different forms of mental imagery.[2]

Questionnaire methods for assessing overall vividness of a subject's mental imagery have been developed for visual imagery (VVIQ)[3], olfactory imagery[4], auditory imagery[5], wine tasting imagery[6], and imagery of movement[7]. As far as I am aware, no questionnaire exists specifically for temperature imagery. Betts' *Questionnaire upon Mental Imagery* included a category for cutaneous imagery which includes prompts for imagining 'the warmth of a tepid bath' and 'the heat of a burning sun'.(Betts, 1909; Sheehan, 1967) The Plymouth Sensory Imagery Questionnaire (Psi-Q) includes prompts for various modalities, including a number that relate to temperature perception: touching warm sand, touching icy water, relaxing in a warm bath, walking briskly in the cold. (Andrade, May, Deeprose, Baugh, & Ganis, 2014)

This use of the concept of vividness in psychology does not necessarily presuppose the differential claim: the practice of measuring the vividness of mental imagery in individuals is consistent with the idea that there is no essential difference in vividness between perceptions and mental images. There does however seem to be a presupposed connection between vividness and perception: the maximum score that subjects in the VVIQ can give to a mental image is 'perfectly clear and as vivid as normal vision'. On the one hand, this

---

2 See: Betts, 1909; D'Angiulli, 2002; Galton, 1880; Marks, 1973, 1995, 1999; Sheehan, 1967.
3 The Vividness of Visual Imagery Questionnaire (VVIQ) (Marks, 1973, 1995, 1999; McKelvie, 1995; Pearson, 1995)
4 The Vividness of Olfactory Imagery Questionnaire (VOIQ). (Gilbert, Crouch, & Kemp, 1998)
5 Clarity of Auditory Imagery Scale (CAIS). (Willander & Baraldi, 2010)
6 Vividness of Wine Imagery Questionnaire (VWIQ). (Croijmans, Speed, Arshamian, & Majid, 2019)
7 Vividness of Haptic Movement Imagery Questionnaire (VHMIQ). (Campos, Lopez, & Perez, 1998)

shows that the scale used considers it a possibility for mental imagery to be as clear as normal vision. On the other hand, this betrays the assumption that perception is as vivid as it gets.

Vividness of mental imagery questionnaires rely on the quality claim *because* individual imagery episodes are evaluated as having a certain degree of vividness.

Kind on vividness

Amy Kind (2017) argues against the use of the concept of vividness in philosophy and psychology, claiming that the notion is too vague and ill-defined to be applicable. Kind considers several candidate notions of visual vividness and evaluates it according to four desiderata: 1) the notion must be coherent, 2) the notion must be applicable to both perception and mental imagery, 3) the notion must typically phenomenally distinguish imaginative experiences from perception, 4) not just any notion that picks out a phenomenal difference will do, it must be something that captures what we mean by *vividness.* If Kind is right in her assessment and the notion of vividness is so ill-defined as to be useless, then the point made in this chapter is moot. If vividness is not a useful concept at all, then arguing for one notion of vividness over the other is pointless. However, I think Kind's paper leaves room for discussion; both because the arguments leave room for counterargument and because she focuses exclusively on visual vividness.

I consider Kind's first desideratum unproblematic and in fact think it is general to scientific concepts. Kind's second desideratum picks out what I call the *quality claim,* and I think it is fair to say that if we want to have a useful notion of vividness, then it must satisfy that desideratum. The third desideratum picks out what I call the differential claim. To have the differential claim as a desideratum for the concept of vividness is strange, because it is possible, as argued above, to use the concept of vividness without endorsing the differential claim. Moreover, if the truth of the differential claim is an empirical matter (and I think it is), then to settle that matter we must be able to use the concept of vividness in way that is neutral to the differential claim. The fourth desideratum is a kind of conceptual conservatism: if the notion of vividness we end up with is very revisionary, then it is not an explication of the concept of vividness, but rather some other concept.

The candidate conceptions of vividness that Kind discusses are *clarity, brightness, color, detail,* and the option of combining such candidates. These candidates pick out properties a visual scene can have, that may contribute to vividness. For the current discussion of vividness in temperature perception, we can ignore clarity, brightness, and color as these do not have an analogue in temperature perception*. Detail,* according to Kind, is best fleshed out as a combination of *determinacy* and *saturation*. Determinacy in Kind's sense pertains to something like the amount of information contained in the mental image. In her example, the image of a striped tiger *simpliciter* is less determinate than the image of a tiger with a certain number of stripes. Saturation pertains to information being present at every point in the visual field. To give an example, a mental image of a tiger with no background is less saturated than a mental image of a tiger in the grass, because it occupies less of the visual field.

Kind rejects *Detail* as a theory of vividness because she claims it does not satisfy the fourth desideratum. She argues that some mental image *A* could feel more vivid than an image *B* even though it is less determinate or saturated. Hence, the proposal of *Detail* does match the explanandum; it is too far removed from what we ordinarily think of as vividness. I find this rejection a little unsatisfactory, so I will return to *determinacy* later in this chapter.

Kind also considers the idea that a combination of the candidate conceptions could provide an analysis of vividness. Of this option she says that it would be "enormously difficult" and that "there is simply not enough content to the notion of vividness to withstand precisification in this way." (Kind, 2017) The bottom line of Kind's discussion is that, in trying to elucidate a vague notion, we either get a notion that is too far from what vividness intuitively means (fails 4), or we get a notion that fails the consistency requirement (1).

One thing I take away from Kinds' discussion is that candidates for the notion of vividness can be properties of the contents of perception (or mental imagery), or combinations of such properties. This view and its negation give us two possible main views of the concept of vividness, which I will discuss in the next two sections.

8.3 The intentional content view

The intentional content view holds that the vividness of an experience depends on the intentional content of the mental state. For example, visual imagery is more vivid if it represents its subject as being more luminous, or as having clear edges. On the intentional content view, if two mental states have a different degree of vividness, they must have different contents.

What I call the intentional content view isn't actually a well-defined view; it's more of a family of views. Different answers can be given as to what makes experiences more vivid, that is *which* content properties contribute to vividness.

The proposals for explaining vividness that Kind considers are examples of the intentional content view. It seems natural on this view to think that vividness can mean something else in a different sense modality: for vision, luminosity may contribute to vividness, but there is no luminosity in auditory experiences.

The answers to the VVIQ also seem to suggest an intentional content view: 'dim' and 'vague' suggest that the difference between vivid mental imagery and non-vivid mental imagery lies in the content of the mental image. It is striking that some of the questionnaires for other modalities use the exact same scale, including terms like 'dim' and 'vague'. These *visual* terms must be meant metaphorically in this non-visual context – which perhaps indicates an implicit rejection of the intentional content view.

The intentional content view tends to emphasize the continuity between perception and imagery. The same content properties (e.g., luminosity) that can make a mental image more or less vivid, can vary in perception. Visual percepts under less-than-ideal lighting conditions can be less vivid, visual percepts of far- or near-sighted people can be less vivid, or we might have less vivid perception when the stimulus is only presented for a short period of time. Still, the proponent of the intentional content can hold the differential claim. The differential claim on the intentional content view would amount to saying that e.g., visual perceptions necessarily have more contrast, luminosity etc. than visual mental imagery.

An intentional content view of vividness might claim vividness is reducible to a single property, or to a combination of properties of the intentional content of the mental state. Any view on which two experiences that have the same content necessarily have the same vividness, is an intentional content view.

8.4 The non-intentional view

The alternative to the intentional content view is that the vividness of an experience does not lie (entirely) in the content. I call this the 'non-intentional view'. If vividness is not part of the *content* of the mental state, then it must be some non-intentional feature of the mental state. On this view, it is possible for two distinct experiences to have the same content but not the same vividness. The appeal of this view lies exactly in that possibility. For example, on this view an episodic memory could have the same content as the original perception but be less vivid. Just like the intentional content view, the non-intentional view is not a well-defined view. It is more of a family of views that share the principle that experiences can differ in vividness without differing in content. There are multiple options as to what exactly vividness is supposed to be if it is not in the content of the mental state.

One way of understanding vividness as a non-intentional of experience is in terms of what Ned Block calls *mental paint.* (Block, 1996, 2003, 2010) Mental paint is those features of experience that are not the intentional content of the experience. The idea behind this is that there's more to experience than its content: according to Block there are mental properties of the experience that *represent* but are not the intentional content of the experience. This is what he calls *mental paint.* To Block, the phenomenal experience of a red tomato represents the tomato being red, but it is not identical to the intentional content *that the tomato is red*. Furthermore, Block thinks there may be mental properties of experience that don't represent. He calls these properties *mental oil.* Block speculates there may be such properties involved in bodily sensations: phenomenal properties of the experience that do not play a role in representation.

Now think of vividness in these terms. The intentional content view I have spelled out above says vividness is a part of the intentional content of experience. But for someone who thinks phenomenal experience is not exhausted by its intentional content, vividness could be a property of the experience without being part of the intentional content. Experiences then can differ in vividness without differing in content. In Block's terms, vividness would be a property that belongs to the category of *mental paint* or *mental oil*.

A mental paint type view of vividness would be committed to vividness lying in the qualitative/phenomenal character of the experience. That commitment, however, is not necessarily a part of a non-intentional view of vividness. There can be versions of the non-intentional view that don't assign an important role to philosophically controversial phenomenal aspects of experience. For example, a non-intentional view could be built on the idea that mental states that are attended to are more vivid than those that are not. Whether a state is attended to or not does not influence the content of the state. Therefore, an attention-based view of vividness would count as a non-intentional view. Attention is often thought of as a gradable notion: there are degrees of attending to mental states. An attention-theory of vividness would be that the vividness of mental states depends on the degree to which they are attended to. This has interesting consequences for vividness, since attention is often thought to be consciously controllable. We can focus our attention on something. So, according to this theory, we ought to be able to make mental images or perceptions more vivid by attending to them.

Consequently, an attention-based view of vividness is unlikely to be combined with endorsement of the differential claim: if attention is consciously controllable, and vividness depends on attention, then we can increase the vividness of a mental state at will. And if the difference between mental imagery and perception is given by the degree of vividness (the differential claim) then it would be possible to turn mental images into perceptions simply by attending to them. This is a bizarre position, so the combination of an attention-based account of vividness and endorsement of the differential claim is a bad idea.

A notion perhaps related to attention is the idea that consciousness is gradable. According to some authors it is possible to be more or less consciously aware of certain mental contents. (Fazekas & Overgaard, 2018; Sandberg & Overgaard, 2015) However, a consciousness-based view of vividness seems ill-conceived. It is perfectly possible to be consciously aware of a non-vivid perception or mental image. As for our unconscious mental states, it is difficult to say whether they are vivid or not: vividness is perhaps a concept only applied to conscious states. Still, it seems wrong to categorize unconscious states prima facie as being non-vivid.

To conclude this section: a non-intentional view can take the form of a qualia-based view or can appeal to more empirically respectable non-intentional features of mental states such as the degree to which they are attended to. The strength of the non-intentional view lies in that it allows for a domain- and modality-general notion of vividness. On the intentional content view, it needs to be specified which content properties count towards vividness, and it is unlikely that this will yield an account that is unified over various sense modalities. A non-intentional account is more likely to give us a notion of vividness under which perceptions from different modalities can be compared.

The downside of a non-intentional account is that the chosen non-intentional feature that underlies vividness may be controversial. Content properties are often straightforwardly characterized and tend to be well-accepted. Non-intentional properties of mental states elicit more resistance. A conception of vividness along the lines of the non-intentional view is perhaps likely to fail Kind’s fourth desideratum, although that would ultimately depend on the specific proposal.

Intermediate views are possible, but the definitions of the views given above would classify them as non-intentional views. One may think that some aspects of the intentional content of experience contribute to vividness, but that vividness is not exhaustively defined by it. I think it is fair to count this as a non-intentional view, because it involves a commitment to some feature of the mental state that is not its intentional content.

8.5 Against the intentional content view

In the following sections I argue that temperature perception and temperature imagery pose a problem for the intentional content view of vividness. To make that argument, I go through three candidate contents that could provide the basis for vividness on the intentional content view: temperature contents (8.5), spatial contents (also 8.5), and determinacy of contents (8.6). I ultimately reject all of them.

Temperature contents
The most natural candidate for a property that is part of the content of temperature perceptions and temperature imagery would be temperature itself. In the previous chapters I have argued for a theory of thermoception that ascribes temperatures to (parts of) the body and to external objects. So, temperatures are part of the intentional content of temperature perceptions (and by extension also of temperature imagery).

The problem with this candidate content is that there would be no distinction between the property represented and the intensity by which it is represented. Take color vision for contrast: we can (to a certain extent) identify colors under a variety of illumination circumstances – illumination and hue are (to a certain degree) perceptually independent. We can say that the property represented (hue) is independent of the intensity by which it is represented (illumination). If we were to give an intentional content view of vividness of color vision, illumination would be a natural candidate for the property that underlies vividness, while hue would not. The content property view would work neatly for color vision precisely the property represented (hue) is independent of the intensity by which it is represented (illumination).

If we choose represented temperature as the property that vividness depends on, then there is no distinction between the property represented and the intensity by which it is represented. It would result in a view that says that representations of very low temperatures and very high temperatures are more vivid, while representations of tepid or comfortable temperatures are less vivid. The problem with this view is that it leads to a position where necessarily experiences of more extreme temperatures are more vivid. Although this is a consistent view, it implies an empirical claim about the phenomenology of temperature experiences and temperature imagery. On this view that temperature vividness depends directly on the temperature represented, two temperature experiences (be they perceptual or imagined) must necessarily have the same level of vividness if they attribute the same temperature. I expect this empirical claim to be false.

To prove this empirical claim false for temperature perception would require an experiment that determines whether the vividness of temperature experiences is (strongly) correlated to the temperature of the stimulus – so strongly that two states that attribute the same temperature always have the same vividness. To show the claim to be false for mental imagery no (physical) experiment is needed: all that is needed is that one can *imagine* two scenarios that attribute the same temperature but differ in vividness. I think this can be done. I can vividly imagine taking a bath at just the right temperature, or I can do so less vividly. Sure, extreme temperatures are more likely to grab one's attention, but that doesn't mean that it is *impossible* to vividly experience a comfortable temperature or non-vividly experience an extreme temperature. When we imagine a temperature experience, it seems plausible that it's possible to imagine it more vividly or less vividly while still imagining the same temperature. This introspective phenomenological methodology is not rigorous, and it could very well be that readers find themselves unable to imagine the same temperature both vividly and less vividly. But the bar for counterevidence to the temperature-content view is low: if it is merely *possible* to imagine the same temperature scenario both vividly and non-vividly, then vividness cannot be fully dependent on temperature content. My contention is that the version of the intentional content view that claims vividness is based on the represented temperature is implausible because it makes a very strong prediction about the phenomenology of temperature perception and imagery that I think will likely be false.

Spatial content

Besides temperature content, temperature experiences can have spatial content. In chapter 6 I distinguished between three modes of temperature perception: body temperature representations, located skin temperature representations and object temperature representations. These modes differ in the spatial contents they attribute. So, could spatial content be the content property that underlies vividness? The view would have to be something like the idea that the stimulus size determines the vividness of the temperature experience, with stimuli that cover a larger area feeling more vivid.

There is a problem with this view: both warmth-responsive afferent fibers and cold-responsive afferent fibers exhibit *spatial summing.* That means that when the area of the stimulus is increased while the temperature of the stimulus is kept constant, the firing rate of these nerve fibers increases. In the case of a warm stimulus, increasing the size of that stimulus makes it feel hotter. In cold fibers, increasing the stimulus area similarly increases activation and results in a *colder* experience, even though the stimulus has maintained a constant level. In many cases, the stimulus is not perceived as being larger, but only as hotter (or colder). (Darian-Smith et al., 1973; Dan R Kenshalo et al., 1967; Schepers & Ringkamp, 2010; Stevens, 1991)

In short, a larger temperature stimulus does not result in a more intense representation of the *same* temperature, it results in an experience that ascribes a *different* temperature to the probe, i.e., an experience with a different temperature content. Spatial content and temperature content are not orthogonal in temperature perception, so the spatial version of the intentional content view would face similar difficulties as the temperature version described above.

Conflation of property and content

The problem with both temperature contents and spatial contents as the basis for vividness is that, in temperature perception and imagery, there is a conflation of represented property and intensity. No independent axes of represented property and intensity exist within the intentional content like hue and illumination are for color vision. A more intense heat stimulus simply *feels hotter*, and a more intense cold stimulus simply *feels colder*. And, while a higher temperature is more intense in a certain sense, it is necessarily more vivid; it seems strange to claim that a vivid temperature experience is simply an *extreme* temperature experience. This conflation of intensity and property makes an intentional content view of vividness implausible.

I believe this conflation of represented property and intensity is extreme in temperature perception, but also present in other sensory modalities. In olfaction, the same odorant can give very different experiences depending on the intensity of the stimulus. A single odorant (skatole) may be experienced as a pleasant smell of flowers, or as a fecal odor depending on the concentration. In audition, it is well-known that frequency content influences perceived loudness, and frequency content is perceived differently at different sound pressure levels. (Fletcher & Munson, 1933) Or, to take a different example, pitch perception can be (slightly) influenced by timbre and loudness of a stimulus. (Oxenham, 2012)

8.6 Determinacy of contents

A third option for the intentional content view is to say that vividness does not depend on the temperature represented or on spatial contents, but on the degree of determinacy of those contents. On this view, vividness would be reduced not to the property represented, but to the determinacy by which that property is represented. This calls back to Kind's discussion of *Detail* as an explanation of vividness. The *determinacy view* of vividness would be something like this: an experience is more vivid if the property represented is represented more determinately.

Using a familiar example, a representation of some object as being red is less determinate than a representation of it as being crimson, because crimson is a determinate of red. Red in turn is a determinable of crimson. The determinate-determinable distinction is a relation between properties. Representations are more determinate when they represent properties that are more determinate. A determinacy view of vividness would claim that the degree of vividness of mental states depends on the determinacy of the properties represented.

For temperature perception and imagery, determinacy could be spelled out as ranges of ranges. The range of temperatures between 10°C and 12°C is a range of the range between 8°C and 20°C. If one represents an object as being of a temperature between 10°C and 12°C that is a more determinate representation than when it is represented as being between 8°C and 20°C. More informal temperature predicates could also be ordered in determinacy, although the use of concepts such as 'warm' or 'hot' is very context dependent: when somebody says they have run you a warm bath, you would not expect it to be boiling hot, even though 'boiling hot' is perhaps a determinate of 'warm'.

There are a few interesting consequences to the determinacy view of vividness. Firstly, a determinacy view of vividness does not provide a basis for the differential claim. It is simply not the case that mental images or memories are always of less determinate (more determinable) properties than perceptions. Secondly, the determinacy view has a consequence for the quality claim. If one takes this route of spelling vividness out in terms of determinacy, and wished to endorse the *quality claim*, then it involves a commitment that all perception and mental imagery that has a certain vividness doesn't just represent something as being F, but it does so with a certain determinacy. This position is indeed taken by Nanay, who thinks all representation comes with a degree of determinacy. (Nanay, 2018, 2020)

There is one big conceptual problem with a determinacy account of vividness. The determinate/determinable distinction is a relation between properties. Strictly speaking we can only call property A more determinate than B if A is a determinate *of B*. But if A is not a determinate or determinable of B then there is no determinacy relation between the two. Crimson and ruby are both determinates of red, but neither is a determinate of the other. There is simply no answer to the question which is more determinate, and there is also no truth to the statement that they are equally determinate. Determinacy is in essence a relational term. Two properties can only be compared in terms of determinacy if such a relation holds, and that relation does not hold between all properties.

This has dire consequences for the quality claim: on a proper interpretation of determinacy, we can only compare the vividness of two mental states if the represented properties stand in a determinate-determinable relation. Say I see something red one moment and see something crimson a moment later. The determinacy theory of vividness predicts the latter episode is more vivid, because crimson is a determinate of red. Now say I see something crimson one moment and something azure a moment later. These two properties do not stand in a determinate-determinable relation to one another, so there is nothing to say about their vividness. This is an odd consequence: vividness isn't usually thought of as an inherently relative term. Something on its own can have a certain vividness, but nothing on its own has a certain determinacy, is my contention. The quality claim was that 'vividness' is a property of a variety of mental states. On the determinacy view, vividness is not an inherent property, but a relational property that only holds between select mental states with contents that stand in a determinate-determinable relation.

Now this problem with determinacy as an explanation for vividness is less acute in the thermoceptive case when we explain determinacy in the way proposed above, as ranges of temperatures. The determinacy of a temperature perception could be reduced to the width of the range of temperatures it represents. Representing something as being between 8°C and 12°C is more determinate than representing something as being between 20°C and 30°C because the range between 8°C and 12°C is narrower. So even though the range 8-12°C is not a sub-range of 20-30°C we can still argue it is more determinate. However, this option isn't open for perceptual representations in general, as not all perceptual properties are structured continuous ranges like the temperature scale. Also, the problem remains when we try to compare the determinacy *between* scales. Sound frequency and temperature are both magnitudes with the right kind of structure, but you can't compare ranges of temperatures to ranges of frequencies. Even within modalities multiple properties can be represented that cannot be compared in this way. On the determinacy proposal, the quality claim can only hold in a very limited sense.

My argument against the intentional content view is never complete, unless I argue against each possible property in the content that it can't be the basis of vividness. Perceptual states and images can be rich in intentional content, so there may be many possible content properties to appeal to as the basis for vividness. However, with important candidates like temperature content, spatial content and the determinacy of content ruled out, I think we have a strong case against the intentional content view. Note that Kind's argument is of the same form: first she lists several prima facie plausible content properties, and then argues that those properties can't be what underlies vividness.

Other content properties of temperature perception besides represented temperature, spatial content, and determinacy, are likely to fail the fourth desideratum. For example, we may think that in perceptually representing a certain very high temperature we also represent it as painful. But to say this painfulness underlies vividness is strange: not all vivid experiences are painful and vice versa. This candidate property fails the fourth desideratum. From the argument about conflation of intensity and property, we can perhaps formulate an extra desideratum for temperature perception specifically: the candidate property that would explain vividness cannot systematically covary with the represented temperature. For if it did, then it would be impossible to have representations of the same temperature with different degrees of vividness.

8.7 Conclusion

In conclusion, I take the intentional content view to be implausible for temperature perception and imagery: it is hard to see what part of the content could count towards vividness. On the non-intentional view, conflation between intensity and quality is not an issue, because this view does not have to pick out a content property on which to base vividness. Consequently, the non-intentional view can accommodate the particulars of temperature perception.

Vividness is understood differently by different authors, and above I have argued that an intentional content view of vividness is not applicable to perception of temperature, while the non-intentional view can accommodate temperature perception. Whether a non-intentional view is a plausible explanation of vividness would depend on the specifics of such a view, and its ability to meet the desiderata discussed above, as well as its applicability in various sensory modalities. I contend that a theory of vividness should be applicable across sense modalities, so theories that do not fit with temperature perception and imagery are at a distinct disadvantage.

If it is true that in thermoception there is no good distinction between intensity of the representation and the property represented, then that can have big implications that lie outside the scope of this chapter. For one, it could have consequences for the possibility of perceptual constancies in temperature perception. Furthermore, it could be an argument against intentionalism; the view that the phenomenal character of perceptual experience supervenes on its intentional content.

Bibliography

Aguggia, M. (2003). Neurophysiology of pain. *Neurological Sciences*, *24*, S57–S60.

Akins, K. (1996). Of sensory systems and the "aboutness" of mental states. *The Journal of Philosophy*, *93*(7), 337–372. https://doi.org/https://doi.org/10.2307/2941125

Andrade, J., May, J., Deeprose, C., Baugh, S. J., & Ganis, G. (2014). Assessing vividness of mental imagery: The plymouth sensory imagery questionnaire. *British Journal of Psychology*, *105*(4), 547–563. https://doi.org/10.1111/bjop.12050

Armstrong, D. M. (1962). *Bodily sensations*. London: Routledge & Kegan Paul.

Armstrong, D. M. (1963). Vesey on sensations of heat. *Australasian Journal of Philosophy*, *41*(3).

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge & Kegan Paul.

Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, *17*, 1016–1031.

Aydede, M. (2009). Is feeling pain the perception of something? *Journal of Philosophy*, *106*(10), 531–567.

Bach-Y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, *221*(5184), 963–964. https://doi.org/10.1038/221963a0

Bain, D. (2017). Evaluativist accounts of pain's unpleasantness. In J. Corns (Ed.), *Routledge Handbook of the Philosophy of Pain* (pp. 9–25).

Bain, D. (2019). Why take painkillers? *Nous*, *53*(2), 462–490. https://doi.org/10.1111/nous.12228

Bautista, D. M., Siemens, J., Glazer, J. M., Tsuruda, P. R., Basbaum, A. I., Stucky, C. L., … Julius, D. (2007). The menthol receptor TRPM8 is the principal detector of environmental cold. *Nature*, *448*(7150), 204–208. https://doi.org/10.1038/nature05910

Bayne, T., & Spener, M. (2010). Introspective humility. *Philosophical Issues*, *20*, 1–22.

Beck, J. (2015). Analogue magnitude representations: A philosophical introduction. *British Journal for the Philosophy of Science*, *66*(4), 829–855. https://doi.org/10.1093/bjps/axu014

Berkeley, G. (1713). *Three dialogues between Hylas and Philonous in opposition to sceptics and atheists*. (C. Choat & A. Haines, Eds.) (Project Gu).

Betts, G. H. (1909). *Distribution and functions of mental imagery (No.26)*. New York: Teachers College, Columbia University.

Bigelow, J., & Pargetter, R. (1987). Functions. *The Journal of Philosophy*, *84*(4), 181–196.

Block, N. (1996). Mental paint and mental latex. *Philosophical Issues*, *7*, 19–49.

Block, N. (2003). Mental paint. In M. Hahn & B. Ramberg (Eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, MA: MIT Press.

Block, N. (2005). Bodily sensations as an obstacle for representationism. In M. Aydede (Ed.), *Pain: New Essays on Its Nature and the Methodology of Its Study* (pp. 137–142).

Block, N. (2010). Attention and mental paint. *Philosophical Issues*, *20*(1), 22–63.

Bourbaki, N. (2004). *Elements of mathematics*. Berlin; Heidelberg: Springer.
Bullock, T. H., & Cowles, R. B. (1952). Physiology of an infrared receptor : The facial pit of pit vipers Author. *Science*, *115*(2994), 541–543.
Cahusac, P. M. B., & Noyce, R. (2007). A pharmacological study of slowly adapting mechanoreceptors responsive to cold thermal stimulation. *Neuroscience*, *148*(2), 489–500. https://doi.org/10.1016/j.neuroscience.2007.06.018
Campero, M., Serra, J., Bostock, H., & Ochoa, J. L. (2001). Slowly conducting afferents activated by innocuous low temperature in human skin. *Journal of Physiology*, *535*(3), 855–865. https://doi.org/10.1111/j.1469-7793.2001.t01-1-00855.x
Campos, A., Lopez, A., & Perez, M. J. (1998). Vividness of visual and haptic imagery of movement. *Perceptual and Motor Skills*, *87*, 271–274.
Casati, R., Dokic, J., & Le Corre, F. (2014). Distinguishing the commonsense senses. In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its Modalities*. https://doi.org/10.1093/acprof
Cervero, F. (2008). Pain theories. In M. C. Bushnell & A. I. Basbaum (Eds.), *The Senses: A Comprehensive Reference vol 5 Pain* (pp. 5–10).
Chuang, H. H., Neuhausser, W. M., & Julius, D. (2004). The super-cooling agent icilin reveals a mechanism of coincidence detection by a temperature-sensitive TRP channel. *Neuron*, *43*(6), 859–869. https://doi.org/10.1016/j.neuron.2004.08.038
Churchland, P. M. (2012). *Plato's camera : how the physical brain captures a landscape of abstract universals*. Cambridge, Mass.: MIT Press.
Coady, C. A. J. (1974). The senses of martians. *Philosophical Review*, *83*(1), 107–125.
Croijmans, I., Speed, L. J., Arshamian, A., & Majid, A. (2019). Measuring multisensory imagery of wine: The vividness of wine imagery questionnaire. *Multisensory Research*, *32*(3), 179–195.
Cummins, R. (1975). Functional analysis. *The Journal of Philosophy*, *72*(20), 741–765.
Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.
Cutter, B., & Tye, M. (2011). Tracking representationalism and the painfulness of pain. *Philosophical Issues*, *21*(1), 90–109.
D'Angiulli, A. (2002). Mental image generation and the contrast sensitivity function. *Cognition*, *85*(1), 11–19. https://doi.org/10.1016/S0010-0277(02)00075-6
Darian-Smith, I., Johnson, K. O., & Dykes, R. (1973). "Cold" fiber population innervating palmar and digital skin of the monkey: responses to cooling pulses. *Journal of Neurophysiology*, *36*(2), 325–346. https://doi.org/10.1152/jn.1973.36.2.325
Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441–458.
Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239
di Pellegrino, G., & Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia*, *66*, 126–133. https://doi.org/10.1016/j.neuropsychologia.2014.11.011
Dowker, F. (2006). Causal sets as discrete spacetime. *Contemporary Physics*, *47*(1), 1–9. https://doi.org/10.1080/17445760500356833
Dretske, F. (1981). Knowledge and the flow of information. Cambridge, MA: MIT Press.
Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, Content, and Function* (pp. 17–36). Oxford: Oxford University Press.

Dubner, R., Sumino, R., & Wood, W. I. (1975). A peripheral "cold" fiber population responsive to innocuous and noxious thermal stimuli applied to monkey's face. *Journal of Neurophysiology*, *38*(6), 1373–1389. https://doi.org/10.1152/jn.1975.38.6.1373
Duclaux, R., & Kenshalo, D. (1972). The temperature sensitivity of the type I slowly adapting mechanoreceptors in cats and monkeys. *Journal of Physiology*, 647–664.
Egeth, H. (1970). The reversal of classical contrast in temperature perception. *Psychonomic Science*, *19*(2), 1970.
Farkas, K. (2008). Phenomenal intentionality without compromise. *Monist*, *91*(2), 273–293. https://doi.org/10.5840/monist20089125
Fazekas, P., & Overgaard, M. (2018). A multi-factor account of degrees of awareness. *Cognitive Science*, *42*(6), 1833–1859. https://doi.org/10.1111/cogs.12478
Filingeri, D., Redortier, B., Hodder, S., & Havenith, G. (2013). The role of decreasing contact temperatures and skin cooling in the perception of skin wetness. *Neuroscience Letters*, *551*, 65–69. https://doi.org/10.1016/j.neulet.2013.07.015
Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *The Bell System Technical Journal*, 377–430. https://doi.org/10.1016/s0016-0032(23)90506-5
Frey, V. (1895). Beitrage zur Sinnesphysiologie der Haut. *Berichten Der Mathematisch-Physischen Classe Der Königl.Sächs. Gesellschaft Der Wissenschaften Zu Leipzig*, *47*, 166–184.
Fulkerson, M. (2011). The unity of haptic touch. *Philosophical Psychology*, *24*(4), 493–516. https://doi.org/10.1080/09515089.2011.556610
Fulkerson, M. (2014). *The first sense: A philosophical study of human touch*. Cambridge, MA: MIT Press.
Gallistel, C. R. (1993). *The organization of learning.* Cambridge, Mass., MIT 1993.
Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain : why cognitive science will transform neuroscience*. Chichester: Wiley-Blackwell.
Galton, F. (1880). Statistics of Mental Imagery. *Mind*, *5*(19), 301–318.
Gerrett, N., Ouzzahra, Y., Redortier, B., Voelcker, T., & Havenith, G. (2015). Female thermal sensitivity to hot and cold during rest and exercise. *Physiology & Behavior*, *152*, 11–19. https://doi.org/10.1016/j.physbeh.2015.08.032
Gilbert, A. N., Crouch, M., & Kemp, S. E. (1998). Olfactory and visual mental imagery. *Journal of Mental Imagery*, *22*(3–4), 137–146.
Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, *28*(3), 344–362.
Goodale, M., & Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8
Goodale, M., & Milner, D. (2013, June 27). Sight unseen: An exploration of conscious and unconscious vision. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199596966.001.0001
Goodman, N. (1972). *Problems and projects*. Indianapolis, New York: Bobs-Merril.
Goodman, N. (1976). *Languages of art : An approach to a theory of symbols*. S.L.: Hackett Publishing Comp.
Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London - Biological Sciences*, *205*(1161), 581–598.

https://doi.org/10.1098/rspb.1979.0086
Gray, R. (2005). On the concept of a sense. *Synthese*, *147*(3), 461–475. https://doi.org/10.1007/s11229-005-1334-1
Gray, R. (2013a). Is there a space of sensory modalities? *Erkenntnis*, *78*(6), 1259–1273. https://doi.org/10.1007/s10670-012-9409-0
Gray, R. (2013b). What do our experiences of heat and cold represent? *Philosophical Studies*, *166*(SUPPL1), 131–151. https://doi.org/10.1007/s11098-012-0083-5
Gray, R. (2014). Pain, perception and the sensory modalities: Revisiting the intensive theory. *Review of Philosophy and Psychology*, *5*(1).
Gray, R. (2023). Thermal Perception and its Relation to Touch. *Philosopher's Imprint*, (Forthcoming).
Grice, H. P. (1962). Some remarks about the senses. In R. J. Butler (Ed.), *Analytical philosophy* (pp. 133–153). Basil Blackwell.
Hall, J. E., & Guyton, A. C. (2011). Guyton and Hall textbook of medical physiology. Philadelphia, PA: Saunders Elsevier. Retrieved from http://www.clinicalkey.com/dura/browse/bookChapter/3-s2.0-C20090602506
Hall, R. J. (2008). If it itches, scratch! *Australasian Journal of Philosophy*, *86*(4), 525–535. https://doi.org/10.1080/00048400802346813
Haugeland, J. (1981). Analog and analog. *Philosophical Topics*, *12*(1), 213–226.
Heil, J. (1983). *Perception and cognition*. Berkeley: Univ. of Calif. Press.
Heil, J. (2011). The senses. In F. Macpherson (Ed.), *The Senses: classic and contemporary philosophical perspectives*. Oxford University Press.
Hensel, H. (1974). Thermoreceptors. *Annual Review of Physiology*, *36*(1), 233–249. https://doi.org/10.1146/annurev.ph.36.030174.001313
Hensel, H. (1982). *Thermal sensations and thermoreceptors in man*.
Hensel, H. (1983). Recent advances in thermoreceptor physiology. *Journal of Thermal Biology*, *8*(1–2), 3–6. https://doi.org/10.1016/0306-4565(83)90063-3
Hensel, H., & Zotterman, Y. (1951a). The effect of menthol on the thermoreceptors. *Acta Physiologica Scandinavica*, *24*(1), 27–34. https://doi.org/10.1111/j.1748-1716.1951.tb00824.x
Hensel, H., & Zotterman, Y. (1951b). The response of the cold receptors to constant cooling. *Acta Physiologica Scandinavica*, *22*(2-3), 96–105. https://doi.org/10.1111/j.1748-1716.1951.tb00758.x
Huang, S. M., Li, X., Yu, Y. Y., Wang, J., & Caterina, M. J. (2011). TRPV3 and TRPV4 ion channels are not major contributors to mouse heat sensation. *Molecular Pain*, *7*, 1–11. https://doi.org/10.1186/1744-8069-7-37
Hume, D. (2010). *David Hume : A treatise of human nature ; a critical edition.* (D. F. Norton & M. J. Norton, Eds.). Oxford: Clarendon Press.
Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: basic minds without content*. Cambridge, Mass: The MIT Press.
Iggo, A., & Muir, A. R. (1969). The structure and function of a slowly adapting touch corpuscle in hairy skin. *The Journal of Physiology*, *200*(3), 763–796. https://doi.org/10.1113/jphysiol.1969.sp008721
Johnson, K. O., Darian-Smith, I., & LaMotte, C. (1973). Peripheral neural determinants of temperature discrimination in man: a correlative study of responses to cooling skin. *Journal of Neurophysiology*, *36*(2), 347–370. https://doi.org/10.1152/jn.1973.36.2.347

Keeley, B. L. (2002). Making sense of the senses : Individuating modalities in humans and other animals. *The Journal of Philosophy*, *99*(1), 5–28.
Kenshalo, D R, & Duclaux, R. (1977). Response characteristics of cutaneous cold receptors in the monkey. *Journal of Neurophysiology*, *40*(2), 319–332. https://doi.org/10.1152/jn.1977.40.2.319
Kenshalo, Dan R, Decker, T., & Hamilton, A. (1967). Spatial summation on the forehead, forearm, and back produced by radiant and conducted heat. *Journal of Comparative and Physiological Psychology*. US: American Psychological Association. https://doi.org/10.1037/h0024610
Kind, A. (2017). Imaginative vividness. *Journal of the American Philosophical Association*, *3*(1), 32–50. https://doi.org/10.1017/apa.2017.10
Klein, C. (2007). An imperative theory of pain. *Journal of Philosophy*, *104*(10), 517–532.
Klein, C. (2015). *What the body commands: The imperative theory of pain*.
Koukou, M. (n.d.). Pain as the evaluation of homeostasis. *Unpublished*.
Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217–230. https://doi.org/10.1038/nrn3008
Kulvicki, J. (2015). Analog representation and the parts principle. *Review of Philosophy and Psychology*, (6), 165–180. https://doi.org/10.1007/s13164-014-0218-z
Kumazawa, T., & Perl, E. R. (1977). Primate cutaneous sensory units with unmyelinated (C) afferent fibers. *Journal of Neurophysiology*, *40*(6), 1325–1338. https://doi.org/10.1152/jn.1977.40.6.1325
LaMotte, R. H., & Thalhammer, J. G. (1982). Response properties of high-threshold cutaneous cold receptors in the primate. *Brain Research*, *244*(2), 279–287. https://doi.org/https://doi.org/10.1016/0006-8993(82)90086-5
Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*, 1439–1459.
Lewis, D. (1971). Analog and digital. *Noûs*, *5*(3), 321–327. https://doi.org/10.2307/2214671
Loar, B. (2017a). Phenomenal intentionality as the basis of mental content. In B. Loar, K. Balog, & S. Beardman (Eds.), *Consciousness and Meaning: Selected Essays*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199673353.003.0017
Loar, B. (2017b). Subjective intentionality. In B. Loar, K. Balog, & S. Beardman (Eds.), *Consciousness and Meaning: Selected Essays*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199673353.003.0011
Locke, J. (1690). *An essay concerning human understanding* (Project Gutenberg).
Long, R. R. (1977). Sensitivity of cutaneous cold fibers to noxious heat: paradoxical cold discharge. *Journal of Neurophysiology*, *40*(3), 489–502. https://doi.org/10.1152/jn.1977.40.3.489
Lumpkin, E. A., & Caterina, M. J. (2007). Mechanisms of sensory transduction in the skin. *Nature*, *445*(7130), 858–865. https://doi.org/10.1038/nature05662
Mackenzie, R. A., Burke, D., Skuse, N. F., & Lethlean, A. K. (1975). Fibre function and perception during cutaneous nerve block. *Journal of Neurology, Neurosurgery and Psychiatry*, *38*(9), 865–873. https://doi.org/10.1136/jnnp.38.9.865
Macpherson, F. (2011a). Individuating the senses. In F. Macpherson (Ed.), *The Senses: classic and contemporary philosophical perspectives* (pp. 3–36). Oxford: Oxford University Press.

Macpherson, F. (2011b). Taxonomising the senses. *Philosophical Studies*, *153*(1), 123–142. https://doi.org/10.1007/s11098-010-9643-8
Macpherson, F. (2014). The space of sensory modalities. In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its Modalities*. Oxford.
Maley, C. J. (2011). Analog and digital, continuous and discrete. *Philosophical Studies*, *155*(1), 117–131. https://doi.org/10.1007/s11098-010-9562-8
Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x
Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery*, *19*(3–4), 153–167.
Marks, D. F. (1999). Consciousness, mental imagery and action. *British Journal of Psychology*, *90*(4), 567–585. https://doi.org/10.1348/000712699161639
Martínez, M. (2011). Imperative content and the painfulness of pain. *Phenomenology and the Cognitive Sciences*, *10*(1), 67–90. https://doi.org/10.1007/s11097-010-9172-0
Martínez, M. (2019). Representations are rate-distortion sweet spots. *Philosophy of Science*, *86*(5), 1214–1226. https://doi.org/10.1086/705493
Matthen, M. (2005). *Seeing, doing, and knowing*. Oxford: Clarendon Press.
Matthen, M. (2015). The individuation of the senses. In M. Matthen (Ed.), *The Oxford Handbook of Philosophy of Perception* (pp. 740–765). Oxford: Oxford University Press.
Matthen, M. (2021). Dual structure of touch: The body versus peripersonal space. In F. de Vignemont (Ed.), *The World at Our Fingertips: A Multidisciplinary Exploration of Peripersonal Space*. Oxford: Oxford University Press.
McGinn, C. (1982). *The character of mind: an introduction to the philosophy of mind*. Oxford: Oxford University Press.
McGinn, C. (1988). Consciousness and content. *Proceedings of the British Academy*, *74*, 225–245.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–836.
McKelvie, S. J. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery*, *19*(3–4), 1–106.
Millikan, R. G. (1996). On swampkinds. *Mind & Language*, *11*(1), 103–117.
Millikan, R. G. (2006a). *On clear and confused ideas : an essay about substance concepts*. Cambridge: Cambridge Univ. Press.
Millikan, R. G. (2006b). *Varieties of meaning : the 2002 Jean Nicod lectures*. Cambridge, Mass.; London: MIT.
Moqrich, A., Hwang, S. W., Earley, T. J., Petrus, M. J., Murray, A. N., Spencer, K. S. R., … Patapoutian, A. (2005). Impaired thermosensation in mice lacking TRPV3, a heat and camphor sensor in the skin. *Science*, *307*(5714), 1468–1472. https://doi.org/10.1126/science.1108609
Müller, J. (1835). *Handbuch der Physiologie des Menschen für Vorlesungen*. J. Hölscher.
Nadel, E. R., Mitchell, J. W., & Stolwijk, J. A. J. (1973). Differential thermal sensitivity in the human skin. *European Journal of Physiology*, *340*(1), 71–76. https://doi.org/10.1007/BF00592198
Nakamura, K. (2011). Central circuitries for body temperature regulation and fever. *American Journal of Physiology - Regulatory Integrative and Comparative*

Physiology, *301*(5). https://doi.org/10.1152/ajpregu.00109.2011
Nanay, B. (2010). A modal theory of function. *The Journal of Philosophy*, *107*(8), 412–431.
Nanay, B. (2012). Action-oriented perception. *European Journal of Philosophy*, *20*(3), 430–446. https://doi.org/10.1111/j.1468-0378.2010.00407.x
Nanay, B. (2013). *Between perception and action*. Oxford: Oxford University Press.
Nanay, B. (2014). Teleosemantics without etiology. *Philosophy of Science*, *81*(5), 798–810. https://doi.org/10.1086/677684
Nanay, B. (2018). Blur and perceptual content. *Analysis*, *78*(2), 254–260. https://doi.org/10.1093/analys/anx119
Nanay, B. (2020). Perceiving indeterminately. *Thought*, *9*(3), 160–166. https://doi.org/10.1002/tht3.454
Neander, K. (1991). Functions as selected effects : The Conceptual Analyst's Defense. *Philosophy of Science*, *58*(2), 168–184.
Neander, K. (1996). Swampman meets swampcow. *Mind and Language*, *11*(1), 118–129. https://doi.org/10.1111/j.1468-0017.1996.tb00036.x
Neander, K. (2017). *A mark of the mental : in defense of informational teleosemantics*.
Neander, K., & Rosenberg, A. (2012). Solving the circularity problem for functions: a response to Nanay. *The Journal of Philosophy*, *109*(10), 613–622.
Nudds, M. (2004). The significance of the senses. *Proceedings of the Aristotelian Society*, *104*(1), 31–51.
O' Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*. https://doi.org/10.1038/381425a0
O'Brien, G. (1998). Connectionism, analogicity and mental content.
O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. *Representation in Mind: New Approaches to Mental Representation*, 1–20. https://doi.org/10.5539/mas.v9n12p118
O'Callaghan, C. (2019). *A multisensory philosophy of perception*. *A multisensory philosophy of perception*. https://doi.org/10.1093/oso/9780198833703.001.0001
O'Dea, J. (2011). A proprioceptive account of sense modalities. In F. Macpherson (Ed.), *The Senses: classic and contemporary philosophical perspectives* (pp. 297–310). Oxford: Oxford University Press.
O'Regan, J. ., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939–1031.
Ouzzahra, Y., Havenith, G., & Redortier, B. (2012). Regional distribution of thermal sensitivity to cold at rest and during mild exercise in males. *Journal of Thermal Biology*, *37*(7), 517–523. https://doi.org/10.1016/j.jtherbio.2012.06.003
Oxenham, A. J. (2012). Pitch perception. *Journal of Neuroscience*, *32*(39), 13335–13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012
Palmer, S. E. (1999). *Vision science: photons to phenomenology.* Cambridge, Mass., MIT 1999.
Papineau, D. (2001). The status of teleosemantics, or How to Stop Worrying about Swampman. *Australasian Journal of Philosophy*, *79*(2), 279–289. https://doi.org/10.1080/713659227
Paricio-Montesinos, R., Schwaller, F., Udhayachandran, A., Rau, F., Walcher, J., Evangelista, R., … Lewin, G. R. (2020). The sensory coding of warm perception. *Neuron*, *106*(5), 830-841.e3. https://doi.org/10.1016/j.neuron.2020.02.035

Peacocke, C. (2019). *The primacy of metaphysics*. Oxford University Press. https://doi.org/10.1093/oso/9780198835578.001.0001
Pearson, D. G. (1995). The VVIQ and cognitive models of imagery: Future directions for research. *Journal of Mental Imagery*, *19*(3–4), 167–170.
Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 74–79. https://doi.org/10.1038/nature12112
Pitt, D. (2004). The phenomenology of cognition or what is it like to think that P, *69*(1), 1–36.
Place, U. T. (2004). *Identifying the mind: selected papers of U.T. Place*. (G. Graham & E. R. Valentine, Eds.). Oxford: Oxford University Press.
Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
Ramsey, W. (2016). Untangling two questions about mental representation. *New Ideas in Psychology*, *40*. https://doi.org/10.1016/j.newideapsych.2015.01.004
Ratcliffe, M. (2012). What is touch? *Australasian Journal of Philosophy*, *90*(3), 413–432. https://doi.org/10.1080/00048402.2011.598173
Romanovsky, A. A. (2007). Thermoregulation: some concepts have changed. Functional architecture of the thermoregulatory system. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, *292*(1), R37-46. https://doi.org/10.1152/ajpregu.00668.2006
Romanovsky, A. A. (2018). The thermoregulation system and how it works. *Handbook of Clinical Neurology*, *156*, 3–43. https://doi.org/10.1016/B978-0-444-63912-7.00001-1
Roxbee Cox, J. W. (1970). Distinguishing the senses. *Mind*, *79*(316), 530–550. Retrieved from http://www.jstor.org/stable/2252438
Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
Ryle, G. (1951). Feelings. *The Philosophical Quarterly*, *1*(3), 193–205.
Sandberg, K., & Overgaard, M. (2015). Using the perceptual awareness scale (PAS). In *Behavioural methods in consciousness research* (pp. 181–195).
Scarry, E. (2006). *Dreaming by the book*. Princeton, N.J.: Princeton University Press.
Schepers, R. J., & Ringkamp, M. (2010). Thermoreceptors and thermosensitive afferents. *Neuroscience and Biobehavioral Reviews*, *34*(2), 177–184. https://doi.org/10.1016/j.neubiorev.2009.10.003
Schonbein, W. (2014). Varieties of analog and digital representation. *Minds and Machines*, *24*(4), 415–438. https://doi.org/10.1007/s11023-014-9342-x
Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, *117*(2), 245–273. https://doi.org/10.1215/00318108-2007-037
Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139173452
Sellars, W. (1963). Philosophy and the scientific image of man. In *Empiricism and the philosophy of mind* (pp. 1–40). London: Routledge & Kegan Paul.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1968.tb00069.x
Shea, N. (2018). Representation in cognitive science. Retrieved from https://openlink.mju.ac.kr/link.n2s?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsoap&AN=edsoap.1001607&amp%0Alang=ko&site=eds-live&scope=site
Sheehan, P. W. (1967). A shortened form of Betts' questionnaire upon mental imagery.

Journal of Clinical Psychology, *23*(3), 386–389. https://doi.org/10.1002/1097-4679(196707)23:3<386::AID-JCLP2270230328>3.0.CO;2-S

Shepard, R. (1978). The mental image. *American Psychologist*, *33*, 125–137.

Shoemaker, S. (2000). Introspection and phenomenal character. *Philosophical Topics*, *28*(2), 247–273.

Skow, B. (2011). Does temperature have a metric structure? *Philosophy of Science*, *78*, 472–489.

Spence, C. (2016). Oral referral: On the mislocalization of odours to the mouth. *Food Quality and Preference*, *50*, 117–128. https://doi.org/10.1016/j.foodqual.2016.02.006

Stevens, J. C. (1979). Thermal intensification of touch sensation: Further extensions of the Weber phenomenon. *Sensory Processes*. US: Academic Press.

Stevens, J. C. (1982). Temperature can sharpen tactile acuity. *Perception & Psychophysics*, *31*(6), 577–580. https://doi.org/10.3758/BF03204192

Stevens, J. C. (1991). Thermal sensibility. In M. A. Heller (Ed.), *The Psychology of Touch*.

Stevens, J. C., & Choo, K. K. (1998). Temperature sensitivity of the body surface over the life span. *Somatosensory and Motor Research*, *15*(1), 13–28. https://doi.org/10.1080/08990229870925

Strang, C. (1960). The perception of heat. *Proceedings of the Aristotelian Society*, *61*, 239–252.

Suárez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, *17*(3), 225–244. https://doi.org/10.1080/0269859032000169442

Tansey, E. A., & Johnson, C. D. (2015). Recent advances in thermoregulation. *Advances in Physiology Education*, *39*(1), 139–148. https://doi.org/10.1152/advan.00126.2014

Tapper, D. N. (1965). Stimulus-response relationships in the cutaneous slowly-adapting mechanoreceptor in hairy skin of the cat. *Experimental Neurology*, *13*(4), 364–385. https://doi.org/10.1016/0014-4886(65)90125-1

Terrien, J., Perret, M., & Aujard, F. (2011). Behavioral thermoregulation in mammals: a review. *Frontiers in Bioscience*, *16*, 1428–1444.

Tewksbury, J. J., & Nabhan, G. P. (2001). Directed deterrence by capsaicin in chillies. *Nature*, *412*(6845), 403–404. https://doi.org/10.1038/35086653

Thaler, L., & Goodale, M. A. (2016). Echolocation in humans: an overview. *Wiley Interdisciplinary Reviews. Cognitive Science*, *7*(6), 382–393. https://doi.org/10.1002/wcs.1408

Thompson, E. (2014). *Waking, dreaming, being : Self and consciousness in neuroscience, meditation, and philosophy*.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Fur Tierpsychologie*, *20*, 410–433.

Tritsch, M. F. (1990). Temperature sensation: The “3-bowls experiment” revisited. *Naturwissenschaften*, *77*(6), 288–289. https://doi.org/10.1007/BF01131227

Tye, M. (2005). Another look at representationalism about pain. In M. Aydede (Ed.), *Pain: New Essays on Its Nature and the Methodology of Its Study* (pp. 99–120). Cambridge MA: MIT Press.

Ungerleider, & Mishkin. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549–586).

van Westen, A. J. (2015). *Powers, manifestations, and their essences*. Utrecht University. Retrieved from http://dspace.library.uu.nl/handle/1874/323958

Vesey, G. N. A. (1960). Berkeley and sensations of heat. *Philosophical Review*, *69*(2).
Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002
Weber, E. H. (1978). *E.H. Weber : the sense of touch*. London; New York: Academic Press for Experimental Psychology Society.
Weiskrantz, L. (1990, August 9). Blindsight: A case study and implications. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198521921.001.0001
Willander, J., & Baraldi, S. (2010). Development of a new clarity of auditory imagery scale. *Behavior Research Methods*, *42*(3), 785–790. https://doi.org/10.3758/BRM.42.3.785
Xu, H., Delling, M., Jun, J. C., & Clapham, D. E. (2006). Oregano, thyme and clove-derived flavors and skin sensitizers activate specific TRP channels. *Nature Neuroscience*, *9*(5), 628–635. https://doi.org/10.1038/nn1692
Yahiro, T., Kataoka, N., Nakamura, Y., & Nakamura, K. (2017). The lateral parabrachial nucleus, but not the thalamus, mediates thermosensory pathways for behavioural thermoregulation. *Scientific Reports*, *7*(1), 5031. https://doi.org/10.1038/s41598-017-05327-8

List of figures