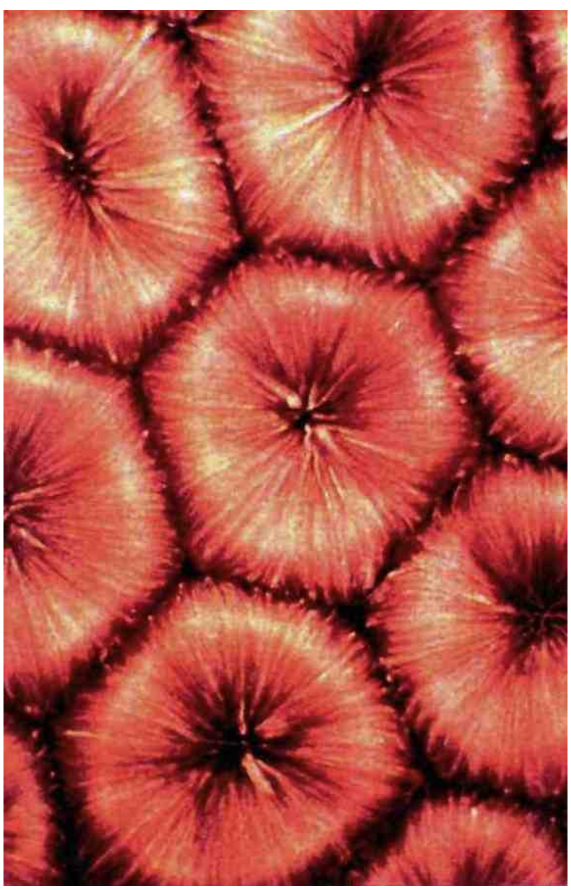
NATIONAL TECHNICAL UNIVERSITY OF ATHENS

Post-graduate Thesis

for the programme

Water Resources Science and Technology



Entropy

Uncertainty in Hydrology and Nature

Nikos Theodoratos

supervised by,

Prof. Demetris Koutsoyiannis

August, 2012

Cover figure

Benard cells. Patterns of convection formed as a liquid is heated from below. Warmer, less dense liquid rises in the center while colder, denser liquid sinks around the edges. The cells are visualized by mixing metal particles in the liquid.

Benard cells are self-organized dissipative systems that arise in the presence of an energy flux, and can potentially be explained by the law of maximum entropy production and the second law of thermodynamics.

Trying to come up with an idea for a cover figure I almost fell victim to the common view that entropy is a measure of "disorder". Thus at first I considered finding a picture of a chaotic phenomenon. In school I was taught that entropy means disorder, a view that was supported by many books and articles that I read while preparing the thesis. But I consider this view to be wrong and misleading; therefore I decided to use an image of Benard cells, which are a manifestation of the second law of thermodynamics despite being self-ordered.

Source: liamscheff.com (http://liamscheff.com/wp-content/uploads/2010/07/convection-cells-closeup.jpg)

Abstract

This thesis presents a review of literature related to the concept of entropy. Its aim is to offer an introduction of the meaning and applications of entropy to a hydrological engineering audience. Furthermore, it discusses some common misconceptions surrounding the concept of entropy, the most important being the identification of entropy with "disorder".

Entropy and related concepts from probability theory, information theory and thermodynamics are defined. In parallel with the definitions, the history of and the technological, scientific and philosophical motivations for the concepts of interest are presented. The main concepts in focus are the second law of thermodynamics; thermodynamic and information-theoretic entropy; entropy production rate and the maximization thereof; statistical inference based on the principle of maximum entropy.

The thesis presents examples of applications of entropy from the area of hydrology, specifically from hydrometeorology and stochastic hydrology; from other natural sciences, specifically from the theory of self-organizing systems with examples of such systems, and from the Earth-System science; and from engineering, specifically an application of measurement system configuration optimization.

The thesis discusses at various sections philosophical questions regarding the concept of entropy.

Περίληψη

Η παρούσα διπλωματική εργασία παρουσιάζει μια επισκόπηση της βιβλιογραφίας σχετικά με την έννοια της εντροπίας. Σκοπός της είναι να προσφέρει μια εισαγωγή της σημασίας και των εφαρμογών της εντροπίας σε υδρολόγους μηχανικούς. Παράλληλα πραγματεύεται διάφορες συνήθεις εσφαλμένες αντιλήψεις σχετικά με την έννοια της εντροπίας, η σημαντικοτέρη εκ των οποίων είναι η ταύτιση της εντροπίας με την «αταξία».

Ορίζονται η εντροπία και σχετικές έννοιες από τη θεωρία των πιθανοτήτων, τη θεωρία της πληροφορίας και τη θερμοδυναμική. Παράλληλα με τους ορισμούς παρουσιάζονται το ιστορικό της ανάπτυξης των εξεταζόμενων εννοιών καθώς και τα τεχνολογικά, επιστημονικά και φιλοσοφικά κίνητρα για την ανάπτυξη αυτή. Οι κύριες έννοιες που εξετάζονται είναι ο δεύτερος νόμος της θερμοδυναμικής, η θερμοδυναμική και πληροφοριοθεωρητική εντροπία, ο ρυθμός παραγωγής εντροπίας και η μεγιστοποίησή του, και η μέθοδος στατιστικής επαγωγής με βάση την αρχή μεγιστοποίησης της εντροπίας.

Η εργασία παρουσιάζει παραδείγματα εφαρμογής της εντροπίας στην Υδρολογία, συγκεκριμένα απο την Υδρομετεωρολογία και τη Στοχαστική Υδρολογία, σε άλλες φυσικές επιστήμες, συγκεκριμένα από τη θεωρία συστημάτων αυτο-οργάνωσης με παραδείγματα τέτοιων συστημάτων, και την Επιστήμη του Γεωσυστήματος, και στην τεχνολογία, συγκεκριμένα μία εφαρμογή βελτιστοποίησης της διάταξης μετρητικών συστημάτων.

Η εργασία πραγματευεται σε διάφορες ενότητες φιλοσοφικά ζητήματα σχετικά με την έννοια της εντροπίας.

Dedicated to

Nikolas Sakellariou

and Nikolas Larentzakis

Table of Contents

Below follows an extended abstract in Greek

Εκτεταμένη Περίληψη

Η παρούσα διπλωματική εργασία παρουσιάζει μια επισκόπηση της βιβλιογραφίας σχετικά με την έννοια της εντροπίας. Το παρόν κεφάλαιο προσφέρει εκτεταμένη περίληψη της εργασίας στα Ελληνικά.

Εισαγωγη

Στο πρώτο κεφάλαιο της εργασίας γίνεται εισαγωγή στην έννοια της εντροπίας. Συγκεκριμένα εξηγείται ότι λίγες επιστημονικές έννοιες είναι τόσο σημαντικές και ταυτόχρονα έχουν τόσο απροσδιόριστη σημασία όσο η εντροπία. Η εντροπία εισήχθη αρχικά για να δώσει μαθηματική μορφή στο δεύτερο νόμο της Θερμοδυναμικής, έχει βρει όμως εφαρμογές σε μεγάλο εύρος επιστημών συμπεριλαμβανομένης της Υδρολογίας. Η παρούσα εργασία απευθύνεται κυρίως σε Υδρολόγους Μηχανικούς. Στόχος της είναι να προσφέρει μία εισαγωγή στην έννοια της εντροπίας, να παρουσιάσει εφαρμογές και συνέπειές της και να ξεκαθαρίσει κάποιες συνήθεις εσφαλμένες αντιλήψεις σχετικά με τη σημασία της.

Στη συνέχεια παρουσιάζεται σύντομη συζήτηση σχετικά με τη σημασία της έννοιας της εντροπίας. Συγκεκριμένα εξηγείται ότι η εντροπία συνήθως σχετίζεται – κατά την άποψή μου λανθασμένα – με την «αταξία» και το «ανακάτεμα». Η άποψη με την οποία συμφωνώ είναι ότι η εντροπία αποτελεί μέτρο της αβεβαιότητας. Η αβεβαιότητα σε κάποιες περιπτώσεις οδηγεί σε μοτίβα, μορφές και «τάξη». Η εικόνα του εξώφυλλου επιλέχθηκε για να δείξει αυτή ακριβώς την άποψη. Δείχνει εξαγωνικές μεταγωγικές κυψέλες, γνωστές ως κυψέλες Benard, οι οποίες, κάτω από κατάλληλες συνθήκες, εμφανίζονται αυθορμήτως όταν ένα λεπτό στρώμα υγρού θερμαίνεται από τα κάτω. Οι κυψέλες εκδηλώνονται λόγω του δεύτερου νόμου της Θερμοδυναμικής, όμως εμφανίζουν δομή και τάξη.

Ετυμολογία, σύντομο ιστορικό και σημασία της εντροπίας.

Η εισαγωγή συνεχίζεται με παρουσίαση της ετυμολογίας της εντροπίας. Η λέξη εντροπία προέρχεται από τα Αρχαία Ελληνικά. Συντίθεται από το πρόθεμα «εν», που σημαίνει «μέσα», «εντός», και από το ουσιαστικό «τροπή», που σημαίνει «στροφή», «μεταμόρφωση», «αλλαγή». Επομένως η κυριολεκτική σημασία της εντροπίας είναι η εσωτερική ή ενδογενής ικανότητα για αλλαγή ή μεταμόρφωση.

Ακολουθεί σύντομο ιστορικό. Ο όρος χρησιμοποιήθηκε επιστημονικά για πρώτη φορά από το Rudolf Clausius το 1863 για να δώσει μαθηματική μορφή στο δεύτερο νόμο της

Θερμοδυναμικής, ο οποίος είχε διατυπωθεί περίπου 15 χρόνια νωρίτερα από τον Clausius (1849) και τον Kelvin (1850) οι οποίοι χρησιμοποίησαν δύο διαφορετικές, αλλά ισοδύναμες, ποιοτικές προτάσεις. Περίπου 30 χρόνια μετά την εισαγωγή της εντροπίας, οι Ludwig Boltzmann και Josiah Gibbs ανέπτυξαν τον κλάδο της Στατιστικής Μηχανικής, στα πλαίσια του οποίου δόθηκε στην εντροπία ένας νέος, πιθανοτικός, ορισμός. 50 χρόνια αργότερα, το 1948, ο ηλεκτρολόγος μηχανικός Claude Shannon, πρωτοπόρος της Θεωρίας της Πληροφορίας, αναζητώντας ένα μέτρο για την αβεβαιότητα της πληροφορίας, ανακάλυψε ότι αυτό το μέτρο θα πρέπει να έχει την ίδια μορφή με την εντροπία κατά Gibbs. Ο Shannon ονόμασε το μέτρο της αβεβαιότητας εντροπία ακολουθώντας προτροπή του Von Neumman. Το 1957 ο Edwin Jaynes, βασισμένος στις εργασίες των Gibbs και Shannon, εισήγαγε την αρχή της μέγιστης εντροπίας, μία μέθοδο στατιστικής επαγωγής η οποία εισήχθη στον κλάδο της Στατιστικής Μηχανικής, αλλά έχει εφαρμογή σε πολλούς κλάδους της επιστήμης. Η έννοια της εντροπίας σχετίζεται με την Υδρολογία κυρίως μέσω της αρχής μέγιστης εντροπίας.

Στη συνέχεια αναφέρονται επιγραμματικά διάφορες ερμηνείες που έχουν δοθεί στην έννοια της εντροπίας. Στις περισσότερες πηγές η εντροπία εξηγείται ως «αταξία», «ανακάτεμα», «χάος», κλπ. Η ερμηνεία αυτή θεωρείται από κάποιους ερευνητές ως λανθασμένη και παραπλανητική, άποψη με την οποία συμφωνώ. Άλλες πηγές ερμηνεύουν την εντροπία ως «μη διαθέσιμη ενέργεια», «χαμένη θερμότητα», κλπ. Αυτές οι ερμηνείες είναι λανθασμένες, αφού η ενέργεια και η εντροπία, αν και σχετίζονται στενά, είναι διαφορετικά φυσικά μεγέθη. Η ερμηνεία την οποία βρίσκω πιο πειστική είναι ότι η εντροπία είναι μέτρο της αβεβαιότητας.

Ορισμοί

Στο δεύτερο κεφάλαιο ορίζονται η εντροπία και άλλες βασικές έννοιες. Οι ορισμοί δε δίνονται σε χρονολογική σειρά. Συγκεκριμένα, πρώτα ορίζεται η εντροπία Shannon και στη συνέχεια η θερμοδυναμική εντροπία. Πιστεύω ότι με αυτή τη σειρά είναι ευκολότερη η κατανόηση της εντροπίας ως μέτρο της αβεβαιότητας.

Θεωρία Πιθανοτήτων

Η εντροπία, τόσο στα πλαίσια της Θεωρίας της Πληροφορίας όσο και στα πλαίσια της θερμοδυναμικής, σχετίζεται πολύ στενά με τη Θεωρία των Πιθανοτήτων. Επομένως η πρώτη ενότητα του κεφαλαίου των ορισμών παρουσιάζει μια σύντομη επανάληψη των βασικών εννοιών της Θεωρίας Πιθανοτήτων. Συγκεκριμένα αναφέρονται οι πρώιμες θεωρίες σχετικά με την τυχαιότητα και τις πιθανότητες, για παράδειγμα οι θεωρίες των Fermat και Pascal. Στη συνέχεια παρουσιάζεται η θεμελίωση των πιθανοτήτων από τον

Kolmogorov με βάση τρία αξιώματα. Ορίζονται οι έννοιες του δειγματικού χώρου, του τυχαίου γεγονότος και του μέτρου πιθανότητας. Επίσης ορίζονται έννοιες όπως το αδύνατο γεγονός, τα αμοιβαίως αποκλειόμενα γεγονότα, τα ανεξάρτητα γεγονότα, οι δεσμευμένες πιθανότητες κλπ. Στη συνέχεια ορίζονται οι τυχαίες μεταβλητές και οι συναρτήσεις κατανομών. Τέλος ορίζεται η έννοια της στοχαστική ανέλιξης καθώς και οι ιδιότητες της στασιμότητας και της εργοδικότητας. Στο Παράρτημα Α γίνεται παρουσίαση επιπλέον ορισμών και ιδιοτήτων από τη θεωρία πιθανοτήτων, για παράδειγμα των αναμενόμενων τιμών και των ροπών.

Θεωρία Πληροφορίας: Εντροπία Shannon

Η δεύτερη ενότητα του κεφαλαίου παρουσιάζει την εντροπία Shannon. Πιο συγκεκριμένα αναφέρεται ότι η Θεωρία της Πληροφορίας αναπτύχτηκε το 1948 από τον Claude Shannon προκειμένου να λύσει σειρά πρακτικών και θεωρητικών προβλημάτων του κλάδου των Τηλεπικοινωνιών. Έκτοτε έχει βρει εφαρμογή σε πολλούς άλλους τεχνολογικούς και ευρύτερα επιστημονικούς κλάδους. Στο πλαίσιο της Θεωρίας της Πληροφορίας η λέξη «πληροφορία» δεν αναφέρεται στο εννοιολογικό περιεχόμενο, δηλαδή τη σημασία, ενός μηνύματος. Αναφέρεται στην αλληλουχία των ψηφίων που αποτελούν το μήνυμα η οποία μπορεί να θεωρηθεί ως μία τυχαία διεργασία.

Κεντρικό ρόλο στη Θεωρία της Πληροφορίας παίζει η έννοια της εντροπίας. Εισήχθη από το Shannon ως μέτρο της αβεβαιότητας. Η αβεβαιότητα μιας τυχαίας μεταβλητής είναι ένα μέγεθος που μπορεί να ποσοτικοποιηθεί ακόμη κι αν η τιμή της μεταβλητής είναι άγνωστη. Για παράδειγμα αν μία μπίλια είναι κρυμμένη σε ένα από τρία κουτιά η αβεβαιότητα για το πού είναι κρυμμένη είναι μικρότερη από ό,τι αν ήταν κρυμμένη σε ένα από εκατό κουτιά.

Παρουσιάζονται αποσπάσματα από την εργασία του Shannon του 1948 όπου υπέθεσε ότι το μέτρο της αβεβαιότητας θα πρέπει να είναι συνάρτηση της κατανομής της τυχαίας μεταβλητής, έστω $H(p_1, p_2,\ldots, p_n)$ και ότι η συνάρτηση αυτή θα πρέπει να ικανοποιεί τρεις ιδιότητες που ακολουθούν την κοινή λογική. Ο Shannon απέδειξε ότι η μοναδική συνάρτηση που ικανοποιεί αυτές τις συνθήκες είναι της μορφής

$$H = -K\sum_{i=1}^{n} p_i \log p_i \qquad \text{(ΕΠ.1)}$$

για οποιοδήποτε θετικό πραγματικό αριθμό K και οποιαδήποτε βάση του λογάριθμου.

Στη συνέχεια δίνεται μία απλή ερμηνεία της εντροπίας Shannon. Συγκεκριμένα, από τη συνάρτηση ΕΠ.1 και για K = 1, προκύπτει ότι η εντροπία ισούται με την αναμενόμενη τιμή του αντίθετου του λογαρίθμου των πιθανοτήτων p_i. Επομένως η εντροπία Shannon μπορεί να ερμηνευθεί ως μια συνάρτηση που καταφέρνει να συνοψίσει μία ολόκληρη κατανομή με έναν αριθμό. Συγκεκριμένα, ο λογάριθμος logp_i μπορεί να ερμηνευτεί ως μέτρο της πιθανοφάνειας του τυχαίου γεγονότος αφού ο λογάριθμος είναι αύξουσα συνάρτηση της πιθανότητας και απεικονίζει το πεδίο τιμών από το [0, 1] στο (-∞ ,0]. Το αντίθετο του λογάριθμου μπορεί να ερμηνευθεί ως μέτρο της απιθανοφάνειας ενός γεγονότος και επομένως η εντροπία ως μέτρο της μέσης απιθανοφάνειας της κατανομής. Μια κατανομή με μεγάλη μέση απιθανοφάνεια, που αντιστοιχεί δηλαδή σε γεγονότα τα οποία είναι κατά μέσο όρο λίγο πιθανά, είναι λογικό να έχει μεγάλη αβεβαιότητα.

Η συλλογιστική του Shannon που οδήγησε στην ανακάλυψη του μέτρου της αβεβαιότητας είναι πολύ ενδιαφέρουσα από επιστημολογική άποψη. Ο Shannon απλά υπέθεσε τρεις ιδιότητες που έπρεπε να έχει η συνάρτηση που έψαχνε με βάση την κοινή λογική και τα μαθηματικά του έδωσαν τη μοναδική συνάρτηση που ικανοποιεί αυτές τις ιδιότητες. Περαιτέρω μελέτη της συνάρτησης έδειξε ότι έχει επιπλέον ενδιαφέρουσες ιδιότητες που δικαιολογούν τη χρήση της ως μέτρο της αβεβαιότητας. Στη συνέχεια της ενότητας της εντροπίας Shannon παρουσιάζονται μερικές από αυτές τις ιδιότητες. Ο τρόπος με τον οποίο ανακαλύφθηκε η εντροπία Shannon δείχνει ότι παρά τις διαφωνίες σχετικά με την ερμηνεία της (βλ. παρακάτω), η συνάρτηση αυτή δεν έχει κάτι το «μαγικό» ή μεταφυσικό. Είναι απλώς η συνάρτηση που έδωσαν τα μαθηματικά όταν ο Shannon ζήτησε τρεις ιδιότητες.

Θερμοδυναμική: Εντροπία Clausius, Boltzmann και Gibbs

Η τρίτη ενότητα του κεφαλαίου των ορισμών πραγματεύεται την εντροπία στα πλαίσια της Θερμοδυναμικής. Η ενότητα αυτή ακολουθεί την ιστορική εξέλιξη των εννοιών της Θερμοδυναμικής. Ξεκινάει επομένως από την Κλασσική Θερμοδυναμική η οποία αναπτύχθηκε παράλληλα με την τεχνολογία των ατμομηχανών και άλλων θερμικών μηχανών. Μεγάλη ήταν η συνεισφορά του Carnot στον κλάδο της Θερμοδυναμικής. Στο πλαίσιο της Κλασσικής Θερμοδυναμικής διατυπώθηκε ο δεύτερος νόμος της Θερμοδυναμικής από τον Clausius το 1849, σύμφωνα με τον οποίο η θερμότητα δε ρέει από ένα ψυχρό σε ένα θερμό σώμα χωρίς να συνοδεύεται από κάποια αλλαγή κάπου αλλού (δηλαδή χωρίς τη δαπάνη έργου), και από τον Kelvin το 1850, σύμφωνα με τον οποίο είναι αδύνατη μια κυκλική διεργασία κατά την οποία θερμότητα μετατρέπεται πλήρως σε έργο. Αποδεικνύεται ότι οι δύο εκφράσεις είναι ισοδύναμες.

Η έννοια της εντροπίας εισήχθη από τον Clausius το 1863 για να δώσει μαθηματική μορφή στο δεύτερο νόμο. Έστω ότι ΔQ είναι μία μικρή ποσότητα θερμότητας που αφαιρείται από ή δίδεται σε ένα σώμα, όπου το ΔQ θεωρείται αρνητικό στην πρώτη περίπτωση και θετικό στη δεύτερη. Έστω επίσης T η απόλυτη θερμοκρασία του σώματος. Τότε η εντροπία ορίζεται ως το πηλίκο:

$$\Delta S = \frac{\Delta Q}{T} \tag{ΕΠ.2}$$

Χρησιμοποιώντας την εντροπία ο δεύτερος νόμος λέει ότι η συνολική εντροπία ενός σώματος και του περιβάλλοντός του αυξάνεται σε κάθε αυθόρμητη διεργασία, δηλαδή:

$$\sum \Delta S \geq 0 \tag{ΕΠ.3}$$

Στη συνέχεια παρουσιάζονται μερικές βασικές συνέπειες του δεύτερου νόμου της Θερμοδυναμικής. Συγκεκριμένα, ένα απομονωμένο σύστημα βρίσκεται σε ισορροπία όταν η εντροπία του είναι μέγιστη. Η ενέργεια έχει τάση προς τη διασπορά[1], δηλαδή σε κάθε διεργασία μία ποσότητα της συνολικής ενέργειας μετατρέπεται σε θερμότητα και δεν είναι δυνατόν η ποσότητα αυτή να μετατραπεί εξ ολοκλήρου σε έργο ή σε άλλη μορφή ενέργειας, επομένως λέγεται ότι η ποιότητα της ενέργειας υποβαθμίζεται. Η τάση διασποράς της ενέργειας υποστηρίζεται ότι θα οδηγήσει στο λεγόμενο θερμικό θάνατο του Σύμπαντος, δηλαδή όλη η ενέργεια του Σύμπαντος θα μετατραπεί σε θερμότητα και το Σύμπαν θα μετατραπεί σε μια ομοιόμορφη «σούπα» θερμότητας στην οποία δε θα συμβαίνει καμία διεργασία. Ο δεύτερος νόμος εισάγει μια ασυμμετρία στη φύση, αφού υπάρχουν μη αντιστρεπτές διεργασίες, για παράδειγμα η αυθόρμητη μεταφορά θερμότητας από ένα θερμό σε ένα ψυχρό σώμα, δεδομένου ότι η αντίστροφη διεργασία δε μπορεί να συμβεί αυθόρμητα. Η ασυμμετρία αυτή δημιουργεί το λεγόμενο βέλος του χρόνου, αφού η παρατήρηση μιας διεργασίας που συμβαίνει σε αντίθεση του δεύτερου νόμου οδηγεί χωρίς αμφιβολία στο συμπέρασμα ότι ο χρόνος κυλάει ανάποδα, όπως για παράδειγμα όταν κάποιος βλέπει μια ταινία ανάποδα.

Στη συνέχεια παρουσιάζεται η συνεισφορά στη Θερμοδυναμική του Maxwell και του Boltzmann. Οι δύο αυτοί μεγάλοι επιστήμονες ανέπτυξαν την Κινητική Θεωρία των Αερίων η οποία επικεντρώνεται στις κατανομές της ταχύτητας και της κινητικής ενέργειας των μορίων ενός αερίου (κατανομές Maxwell-Boltzmann) αντί να προσπαθήσει να περιγράψει τις εξισώσεις κινήσεις του καθενός μορίου. Αυτή η οπτική γωνία οδήγησε στην καθιέρωση του κλάδου της Στατιστικής Μηχανικής και αποτέλεσε σημείο καμπής στη γενικότερη εξέλιξη της σύγχρονης Φυσικής.

---

[1] Ο όρος «διασπορά» χρησιμοποιείται για να αποδώσει τον αγγλικό όρο “dissipation”. Άλλοι όροι, που έχουν χρησιμοποιηθεί για την ίδια έννοια είναι: «σκέδαση», «διάχυση», «διασκόρπιση».

Ο Boltzmann κατάφερε να αποδείξει ότι ένα αέριο σε ισορροπία υπακούει στην κατανομή Maxwell-Boltzmann ανεξαρτήτως της αρχικής κατανομής ταχυτήτων. Επομένως συνειδητοποίησε ότι η κατανομή της ταχύτητας πρέπει να συνδέεται με την εντροπία και κατέληξε σε καινούριο ορισμό της εντροπίας. Κατ' αρχάς όρισε την έννοια της μακροκατάστασης, η όποια είναι η κατάσταση ενός συστήματος όπως παρατηρείται μακροσκοπικά και περιγράφεται από μεταβλητές όπως ο όγκος, η πίεση, και η θερμοκρασίας, και την έννοια της μικροκατάστασης, η οποία είναι η διάταξη των θέσεων και ταχυτήτων όλων των σωματιδίων ενός συστήματος. Κάθε μακροκατάσταση μπορεί να προκύπτει από τεράστιο πλήθος μικροκαταστάσεων. Επομένως ένα σύστημα μπορεί να βρίσκεται σε μακροσκοπική ισορροπία, αλλά η μικροκατάσταση στην οποία βρίσκεται συνεχώς αλλάζει. Ο Boltzmann υπέθεσε ότι η μακροκατάσταση ισορροπίας προκύπτει από W τω πλήθος ισοπίθανες μικροκαταστάσεις και απέδειξε ότι η εντροπία στην κατάσταση ισορροπίας ορίζεται ως:

$$S = k \ln W = - k \ln p_i \qquad \text{(ΕΠ.4)}$$

όπου $p_i = \frac{1}{W}$ η πιθανότητα να βρίσκεται το σύστημα σε κάποια μικροκατάσταση i και k η σταθερά Boltzmann. Η υπόθεση ότι οι μικροκαταστάσεις ισορροπίας είναι ισοπίθανες ονομάζεται εργοδική υπόθεση.

Στη συνέχεια παρουσιάζονται νέες ερμηνείες του δεύτερου νόμου της Θερμοδυναμικής και της έννοιας της εντροπίας στις οποίες οδήγησε ο νέος ορισμός της εντροπίας από το Boltzmann. Κατ' αρχάς δόθηκε στατιστική ερμηνεία στην εντροπία. Επιπλέον, δεδομένου ότι η κατάσταση ισορροπίας αντιστοιχεί στη μέγιστη εντροπία, η ισορροπία μπορεί να οριστεί ως η μακροκατάσταση εκείνη που προκύπτει από το μέγιστο αριθμό μικροκαταστάσεων. Σχετικά με αυτό το συμπέρασμα η εργασία προσφέρει το παράδειγμα της διαστολής απομονωμένου αερίου και υποστηρίζει ότι τέτοιου είδους παραδείγματα έχουν συμβάλει στην άποψη ότι η εντροπία είναι μέτρο της «αταξίας».

Ακολουθεί παρουσίαση φιλοσοφικών ζητημάτων που προέκυψαν από τις συνεισφορές του Boltzmann. Συγκεκριμένα οι κυρίαρχες θρησκευτικές και φιλοσοφικές αντιλήψεις της εποχής δε μπορούσαν να συμβαδίσουν με το δεύτερο νόμο της Θερμοδυναμικής, ιδιαίτερα το στατιστικό ορισμό της εντροπίας του Boltzmann. Από πολλούς απορρίφθηκε εντελώς ο δεύτερος νόμος. Άλλοι, όπως ο Maxwell, προσπάθησαν να τον ερμηνέυσουν ως ανθρώπινη ψευδαίσθηση χρησιμοποιώντας ένα πείραμα σκέψης, το λεγόμενο Δαίμονα του Maxwell, που παρουσιάζεται στην εργασία. Σε απάντηση ο Boltzmann διατύπωσε το Η-Θεώρημα. Αν και το θεώρημα αυτό θεωρείται στις μέρες μας λανθασμένο, οι βασικές θέσεις του Boltzmann σε σχέση με την Ατομική Θεωρία της Ύλης έχουν επιβεβαιωθεί. Ο Boltzmann δεν κατάφερε όμως να επηρεάσει τις κυρίαρχες αντιλήψεις της εποχής του και εικάζεται ότι αυτό το γεγονός ήταν ίσως ένας από τους λόγους που τον ώθησαν στην αυτοκτονία του.

Το ερώτημα γιατί ισχύει ο δεύτερος νόμος της Θερμοδυναμικής παραμένει αναπάντητο. Στην εργασία αναφέρω ότι η άποψη με την οποία συμφωνώ είναι ότι η ισχύς του νόμου οφείλεται στη στατιστική του φύση. Συγκεκριμένα ο δεύτερος νόμος ισχύει γιατί η πιθανότητα να αυξηθεί η εντροπία ενός συστήματος είναι συντριπτικώς μεγαλύτερη από την πιθανότητα να μειωθεί. Όμως δεν υπάρχει κάποια αρχή της φυσικής που να μην επιτρέπει να μειωθεί η εντροπία, αν και κάτι τέτοιο είναι εξαιρετικά απίθανο.

Στη συνέχεια παρουσιάζεται η εντροπία Gibbs η οποία αποτελεί βελτίωση της εντροπίας Boltzmann. Συγκεκριμένα, η εντροπία Gibbs ορίζεται και σε συστήματα εκτός ισορροπίας. Σε τέτοια συστήματα οι μικροκαταστάσεις δεν είναι ισοπίθανες. Επομένως ο Gibbs αντί να ορίσει την εντροπία ως συνάρτηση του λογαρίθμου $\log p_i$ την ορίζει ως συνάρτηση της αναμενόμενης τιμής αυτού του λογάριθμου, δηλαδή:

$$S = -k \sum_i p_i \ln p_i \tag{ΕΠ.5}$$

Η ενότητα της Θερμοδυναμικής κλείνει με την επιγραμματική παρουσίαση του ρυθμού παραγωγής εντροπίας, ο οποίος αποτελεί βασικό μέγεθος της Θερμοδυναμικής συστημάτων εκτός ισορροπίας. Επίσης αναφέρονται η αρχή ελαχιστοποίησης του ρυθμού παραγωγής εντροπίας του Prigogine και η αρχή μεγιστοποίησης του ρυθμού παραγωγής εντροπίας του Ziegler. Εξηγείται ότι οι δύο αυτές αρχές δεν είναι αντιφατικές, αφού η πρώτη αναφέρεται σε μη μόνιμες διεργασίες κοντά στην ισορροπία ενώ η δεύτερη σε μόνιμες διεργασίες μακριά από την ισορροπία. Σύμφωνα με την αρχή του Ziegler ένα σύστημα εκτός ισορροπίας προσπαθεί να τείνει προς την ισορροπία με το γρηγορότερο δυνατό ρυθμό. Οι επιπτώσεις αυτής της αρχής είναι σημαντικές. Κάποιες από αυτές παρουσιάζονται στο τρίτο κεφάλαιο.

Αρχή μέγιστης εντροπίας

Η τέταρτη ενότητα του κεφαλαίου των ορισμών παρουσιάζει την αρχή μέγιστης εντροπίας, μία μέθοδο στατιστικής επαγωγής που χρησιμοποιεί την εντροπία Shannon και εισήχθη από τον Jaynes το 1957. Η μέθοδος εισήχθη στον κλάδο της Στατιστικής Μηχανικής προκειμένου να υπολογιστούν κατανομές και εξισώσεις χρησιμοποιώντας μόνο βασικούς νόμους της Φυσικής χωρίς επιπλέον υποθέσεις όπως η εργοδική υπόθεση. Πλέον εφαρμόζεται ως γενική μέθοδος στατιστικής επαγωγής σε πολλούς επιστημονικούς κλάδους.

Η μέθοδος υπολογίζει την κατανομή p_i που μεγιστοποιεί την εντροπία $H(p_i)$, δηλαδή που μεγιστοποιεί την αβεβαιότητα, χρησιμοποιώντας διαθέσιμες πληροφορίες, για παράδειγμα πειραματικές μετρήσεις, ως περιορισμούς. Με αυτόν τον τρόπο λαμβάνονται υπ' όψιν όλα τα γνωστά δεδομένα, αλλά δε χρησιμοποιούνται επιπλέον δεδομένα που δεν είναι διαθέσιμα, για παράδειγμα υποθέσεις για τον τύπο της κατανομής. Σύμφωνα με τον Jaynes (1957) «*η αρχή μέγιστης εντροπίας δεν είναι εφαρμογή κάποιου φυσικού νόμου αλλά απλώς μία συλλογιστική μέθοδος που εξασφαλίζει ότι δεν εισάγεται ασυνείδητα κάποια αυθαίρετη υπόθεση*».

Στη συνέχεια παρουσιάζονται παραδείγματα υπολογισμού διαφόρων κατανομών για διάφορους περιορισμού. Για παράδειγμα όταν δεν υπάρχει κανένα δεδομένο, δηλαδή κανένας περιορισμός, η μέθοδος καταλήγει στην ομοιόμορφη κατανομή[2], αποδεικνύοντας γιατί η πιθανότητα κάθε ζαριάς είναι 1/6.

Σημειώνεται ότι η εργασία του Jaynes του 1957 περιέχει πολλές ενδιαφέρουσες φιλοσοφικές παρατηρήσεις του Jaynes κάποιες εκ των οποίων παρουσιάζονται στο Παράρτημα Γ.

Θερμοδυναμική και πληροφοριοθεωρητική εντροπία

Η πέμπτη ενότητα του κεφαλαίου παρουσιάζει πλευρές της συζήτησης σχετικά με τη σχέση ή μη μεταξύ της θερμοδυναμικής εντροπίας (Clausius, Boltzmann, Gibbs) και της πληροφοριοθεωρητικής εντροπίας (Shannon). Η συζήτηση αυτή είναι ακόμη ανοιχτή και έχει φιλοσοφικές προεκτάσεις. Μία άποψη είναι ότι οι δύο εντροπίες, μολονότι δίνονται από την ίδια εξίσωση, δε σχετίζονται. Μία άλλη άποψη, με την οποία συμφωνώ, είναι ότι οι δύο εντροπίες σχετίζονται στενά και μάλιστα ότι η θερμοδυναμική εντροπία αποτελεί ειδική περίπτωση της εντροπίας Shannon.

Στην εργασία παρουσιάζονται δύο λόγοι οι όποιοι φαίνεται να οδηγούν στην πρώτη άποψη. Ο πρώτος είναι ότι η εντροπία Shannon είναι αδιάστατο μέγεθος ενώ η θερμοδυναμική εντροπία έχει μονάδες J/K. Εξηγείται όμως ότι αυτό οφείλεται στο ότι η θερμοκρασία έχει μονάδες K για ιστορικούς λόγους, ενώ θα μπορούσε να έχει μονάδες 1/J οι οποίες θα καθιστούσαν τη θερμοδυναμική εντροπία αδιάστατη. Ο δεύτερος λόγος είναι ότι η λέξη «πληροφορία» παρερμηνεύεται ως κάτι το υποκειμενικό, με την έννοια ότι εξαρτάται από τη γνώση κάποιας οντότητας.

[2] Η απόδειξη βρίσκεται στο Παράρτημα Β.

Εφαρμογές

Στο τρίτο κεφάλαιο παρουσιάζονται εφαρμογές της εντροπίας στον κλάδο της Υδρολογίας, ευρύτερα στις Φυσικές Επιστήμες και στην Τεχνολογία.

Η παρουσίαση ξεκινάει από τον κλάδο της Υδρομετεωρολογίας. Ο κλάδος αυτός ασχολείται με τη θερμοδυναμική της ατμόσφαιρας. Επομένως η έννοια της εντροπίας απαντάται συχνά στην Υδρομεωρολογία. Αναφέρεται το παράδειγμα του τεφιγράμματος. Στη συνέχεια παρουσιάζονται εφαρμογές από τον κλάδο της Στοχαστικής Υδρολογίας. Συνήθης εφαρμογή της εντροπίας σε αυτό τον κλάδο είναι η αρχή μέγιστης εντροπίας με τη βοήθεια της οποίας υπολογίζονται κατανομές υδρολογικών μεταβλητών. Άλλες εφαρμογές ασχολούνται με τη γεωμορφολογία και τη χρονική κατανομή της βροχής. Τέλος η εντροπία χρησιμοποιείται για την εξήγηση της διεργασίας Hurst-Kolmogorov και της ιδιότητας της εμμονής, γνωστής και ως μακροπρόθεσμης «μνήμης».

Οι εφαρμογές της εντροπίας στις Φυσικές Επιστήμες είναι πολυπληθείς. Οι εφαρμογές που παρουσιάζονται στην εργασία επιλέχθηκαν πρώτον για να προκαλέσουν ενδιαφέρον για την εντροπία και δεύτερον για να δείξουν ότι η ταύτιση της εντροπίας με την «αταξία» μπορεί να είναι παραπλανητική. Οι έννοιες αυτής της ενότητας παρουσιάζονται με εμπειρικό τρόπο και κάποιες από τις αναλύσεις βασίζονται στη διαίσθησή μου και όχι σε επιστημονικές αποδείξεις.

Στο πρώτο μέρος επιχειρείται να συνδεθεί η αρχή μεγιστοποίησης της ρυθμού παραγωγής εντροπίας του Ziegler με φαινόμενα «αυτο-οργάνωσης». Η λεγόμενη αυτο-οργάνωση εμφανίζεται σε συστήματα που διατηρούνται εκτός ισορροπίας από ροή ενέργειας ή από την ύπαρξη βαθμίδων (διαφορά πίεσης, συγκέντρωσης, θερμοκρασίας, κλπ). Τείνοντας προς την ισορροπία παράγουν εντροπία. Η αυτο-οργάνωση πιθανώς μεγιστοποιεί το ρυθμό παραγωγής εντροπίας. Παρουσιάζονται παραδείγματα όπως οι μεταγωγικές κυψέλες Benard, η τύρβη, η ώσμωση διάμεσου λιπιδικών μεμβρανών. Παρουσιάζεται η άποψη ότι οι ζωντανοί οργανισμοί αποτελούν αυτο-οργανωμένα συστήματα. Επίσης εξηγείται ότι το φαινόμενο της ζωής, αν και οδηγεί σε όλο και περισσότερη οργάνωση και «τάξη» δεν αντιβαίνει το δεύτερο νόμο της Θερμοδυναμικής. Το μέρος αυτό κλείνει με την εκτίμηση ότι αν η ζωή όντως αποτελεί απόρροια της μεγιστοποίησης της παραγωγής εντροπίας, τότε η εμφάνισή της είναι πολύ πιθανή όταν υπάρχουν κατάλληλες συνθήκες. Στο δεύτερο μέρος παρουσιάζεται η έννοια του Γεωσυστήματος, το οποίο αποτελεί συνδυασμό της βιόσφαιρας, της ατμόσφαιρας, της υδρόσφαιρας και της λιθόσφαιρας σε ένα ενιαίο σύστημα. Εξηγείται ότι η έννοια της εντροπίας έπαιξε κεντρικό ρόλο στην εμφάνιση και εξέλιξη της έννοιας του Γεωσυστήματος.

Παρουσιάζεται ένα παράδειγμα τεχνολογικής εφαρμογής της εντροπίας. Συγκεκριμένα παρουσιάζεται πώς η αρχή μέγιστης εντροπίας μπορεί να βοηθήσει στο σχεδιασμό μετρητικών συστημάτων και στην επιλογή των βέλτιστων θέσεων των αισθητήρων.

Επίλογος

Η εργασία κλείνει με τον επίλογο ο οποίος ανακεφαλαιώνει τα βασικά σημεία της. Συγκεκριμένα επαναλαμβάνεται η άποψη ότι η ερμηνεία της εντροπίας ως μέτρο της «αταξίας» είναι παραπλανητική και ότι η ερμηνεία της ως μέτρο της αβεβαιότητας είναι πιο σωστή. Επίσης επαναλαμβάνεται η άποψη ότι η θερμοδυναμική και η πληροφοριοθεωρητική εντροπία είναι στενά σχετιζόμενες έννοιες, με την πρώτη να αποτελεί ειδική περίπτωση της δεύτερης. Στη συνέχεια απαριθμούνται οι εφαρμογές που παρουσιάστηκαν και ξεκαθαρίζεται ότι πολλά ζητήματα σχετικά με την εντροπία παραμένουν ανοιχτά.

Preface

Entropy is one of the most important concepts of science, and, at the same time one of the most confusing and misunderstood, for students and scientists alike. Originating in the field of thermodynamics, it is used to explain why certain processes seem to be able to occur only in one direction. The field of application of entropy however is much wider, almost universal. It is not an exaggeration to say that entropy can be used to explain processes taking place in the hearts of the farthest stars and in the nuclei of the smallest living cells.

My goal in writing this thesis is to give a comprehensive introduction to the concept of entropy. A single thesis cannot cover the whole topic. Therefore this document aims at showing the breadth of potential applications in hydrology and beyond, at discussing some common misconceptions about entropy, and, hopefully at helping readers to grasp its essence.

In writing this thesis I hope that I will offer something back to the *Department of Water Resources and Environmental Engineering* of the *National Technical University of Athens* for offering me the opportunity to attend the interdisciplinary post-graduate program *Water Resources Science and Technology*; and to the *ITIA* research group for supervising my studies and research since 2003. I hope that this thesis will help future students to appreciate and see beauty in entropy, and perhaps gain some better understanding in the concepts surrounding it.

Entropy is a measure of uncertainty. Perhaps this is why it is so difficult to grasp, as our human mind is afraid of losing control. Therefore it prefers to pretend that there is some certainty. However it is uncertainty that gives rise to beauty both in our lives and in Nature, as I show with some examples in this thesis.

I feel very grateful towards Professor Demetris Koutsoyiannis for suggesting such a beautiful topic for my thesis and for having been directly or indirectly a guide and inspiration for my scientific and academic development. I would like to thank Professor Constantinos Noutsopoulos and Professor Nikos Mamasis, members of my examination committee for their critical comments which helped me improve this document.

I would like to thank Kamila for offering me support, advice and some extremely interesting examples of entropy during the preparation of the thesis. I would also like to thank Theodora, Marcel and Diogo for discussions regarding entropy. Finally I would like to thank and dedicate this thesis to my friend Nikolas Sakellariou for inspiring me in the spring of 2010 to leave my consulting job in San Francisco and return to Academia and his uncle Nikolas Larentzakis for offering me support in the first few months upon my return to Athens.

27/08/2012

1. Introduction

Very few scientific concepts are as universally important, and, at the same time, as elusive, as *entropy*[3]. It was first introduced to give a mathematical form to the *second law of thermodynamics*. However it has found applications in a wide range of sciences, including Hydrology. This thesis is directed mainly to a hydrologic engineering audience. Its goal is to introduce the concept of entropy, to present some of its implications and applications, and to clarify some common misconceptions about its meaning.

Entropy is usually associated – in my view wrongly – with "*disorder*" and "*mixed-upness*". My point of view is that entropy is a measure of *uncertainty*, which sometimes leads to patterns, forms and "order". The cover image of the thesis was chosen to demonstrate exactly this fact. It shows hexagonal convection cells, known as Benard cells, which appear spontaneously, under the correct conditions, when a thin layer of liquid is heated from below. They emerge due to the second law of thermodynamics, yet they exhibit structure and order.

Etymology of entropy

The word entropy originates from Greek, where it is called *εντροπία.* It is synthesized by the prefix "*εν*", which means "in", and the noun "*τροπή*", which means a "turning", a "transformation", a "change". Therefore entropy literally means the internal capacity for change or transformation. The term was firstly used in a scientific context by Rudolf Clausius in 1863 (Ben-Naim, 2008).

He used an Ancient Greek term because he preferred "*going to the ancient languages for the names of important scientific quantities, so that they mean the same thing in all living tongues*" (*Ibid.*).

History of entropy[4]

Entropy was introduced in 1863 within the field of thermodynamics to give a mathematical expression to the second law of thermodynamics. The law was first formulated around 15 years earlier, in 1849, by Clausius as a qualitative statement which, paraphrased, states

[3] In this document I use italics to give emphasis to words or phrases. Short quotes from other works are presented between quotation marks and in italics, while longer ones are formatted as block quotations, i.e. indented and in italics.

[4] A more detailed account of the history of thermodynamics, entropy and the second law is presented in Chapter 2.

that *heat does not flow spontaneously from a cold to a hot body*. A year later Kelvin formulated it as a different, but equivalent, qualitative statement, which, paraphrased, states *that no engine can turn 100% of the used energy to work*. Using the concept of entropy Clausius introduced a new formulation, which states that *the entropy of a system and its surroundings increases during a spontaneous change*, from which it results that *the entropy of a system at equilibrium is maximum*.

In coining the term "entropy", Clausius chose to use the noun "τροπή" to denote transformation and the prefix "εν" to make it sound like the word *energy* because "*these two quantities are so analogous in their physical significance, that an analogy of denominations seems to be helpful*" (Ben-Naim, 2008).

Around 30 years later Ludwig Boltzmann and Josiah Gibbs developed the field of *statistical mechanics* which explains the properties and laws of thermodynamics from a statistical point of view, given that even a small quantity of matter is composed of an extremely large number of particles. Within statistical mechanics entropy was given a new, probabilistic, definition. It was this definition that has led to the widespread but misleading view that entropy is a measure of disorder.

Another 50 years later, in 1948, Claude Shannon, an electrical engineer, pioneered the field of information theory. In his search for a measure of informational uncertainty he discovered that such a measure should have the same mathematical form as Gibbs' entropy. So Shannon gave to his measure the name entropy. He chose this name following the suggestion of John Von Neumann. According to Tribus and McIrvine (1971) Shannon was greatly concerned how to call his measure. Von Neumann told him

> *You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.*

The fact that the thermodynamic and information-theoretic entropies have the same formula and name has led to a debate about whether the two concepts are identical or not. My view is that they are the same concept. Both entropies are a measure of uncertainty, just applied to two different phenomena, namely to the random moves of particles on the one hand, and to the transmission of information on the other.

Based on the works of Gibbs and Shannon, Edwin Jaynes introduced in 1957 the principle of maximum entropy, which is a method of statistical inference. Jaynes introduced it within the field of statistical mechanics. But as a method of inference, it has applications in many fields. It is mostly through this principle that the concept of entropy is related to Hydrology.

Meaning of entropy

In most textbooks it is explained that entropy means "*disorder*", "*mixed-upness*", "*chaos*", etc. (e.g. Cengel and Boles, 2010). I find these explanations misleading. Disorder and mixed-upness are subjective concepts. They do capture some features of entropy and the second law of thermodynamics. But they are misleading because some of the manifestations of the law, such as the Benard cells, lead to the opposite of what is usually meant as disorder.

Other sources explain that entropy is "*unavailable energy*", "*lost heat*", etc. (e.g. Dugdale, 1996). These explanations are wrong. Energy and entropy are different physical quantities with completely different meanings. A form of energy ("unavailable" or "heat") cannot be used to explain entropy. As Ben-Naim (2008) explains

> *both the "heat loss" and "unavailable energy" may be applied under certain conditions to TΔS but not to entropy. The reason it is applied to S rather than to TΔS, is that S, as presently defined, contains the units of energy and temperature. This is unfortunate. If entropy had been recognized from the outset as a measure of information, or of uncertainty, then it would be dimensionless, and the burden of carrying the units of energy would be transferred to the temperature T.*

Another point of view, which I find convincing, is that entropy is a measure of *uncertainty*. This point of view will become much clearer in the following chapters. The fact that uncertainty can be measured may sound surprising. But even if the outcome of a phenomenon is uncertain, the degree of uncertainty can be quantified. For example if a ball is hidden in one out of three boxes the uncertainty is much less than if it were hidden in one out of one hundred boxes. Or, for example, the uncertainty for the highest temperature on a random spring day, say March 17th, in a tropical country is much less than the uncertainty in a Mediterranean country such as Greece.

Ben-Naim (2008) suggests that entropy is a measure of "*missing information*". The word "information" is potentially misleading. As it is explained in Chapter 2 the meaning of the word in the context of information theory does not refer to the content of a message but to the message itself, more specifically to the letters that are used to construct a message. Entropy measures the uncertainty of selecting one or another letter, if a message is considered as a stochastic series of letters. With this clarification in mind, the terms "uncertainty" and "missing information" are equivalent.

It should be clarified here that uncertainty should not be viewed as something subjective. Uncertainty does not exist because *we* are not certain about the outcome of a phenomenon. Similarly missing information does not refer to the fact that *we* do not know the outcome. Uncertainty, or missing information, is a property of the phenomenon *itself*.

Outline of the thesis

The concept of entropy along with related concepts from probability theory, information theory and thermodynamics, will be defined in detail in Chapter 2. Chapter 3 is dedicated to applications of entropy and related concepts to the field of hydrology and beyond. The thesis concludes with the epilogue in Chapter 4. The concept of entropy, both historically when it was being introduced, and in the present as new implications and applications are being discovered, pushes the limits of our fundamental understanding of nature. As such it is a very "philosophical" concept and therefore the thesis discusses at various sections philosophical questions that are related to entropy.

2. Definitions

This chapter provides definitions of the main concepts that are related to entropy from the areas of probability theory, statistics, information theory, and thermodynamics. It also describes briefly how the concepts developed historically and what technological, scientific and philosophical questions they came to answer.

The concept of entropy, whether in the thermodynamic or information-theoretic sense, is closely related to probability theory and statistics. Therefore the first section of the chapter is dedicated to the introduction of the main concepts of probabilities. Some detailed definitions are presented in Appendices A and B.

The next two sections present entropy first from an information-theory and then from a thermodynamic perspective. They do not follow the order that the concepts were historically developed. I hope that using a reverse order the concepts will be easier to follow.

The fourth section presents the principle of maximum entropy, which is a method of statistical inference.

Finally, the fifth section presents the debate about the relation between information-theory and thermodynamic entropy.

2.1 Probability Theory

Early theories of probability

Probability theory is the branch of mathematics that studies probabilities. It was originally developed in the 16th century by the Italian Renaissance mathematician, astrologer and gambler *Gerolamo Cardano* and in the 17th century by the French mathematicians *Pierre de Fermat* and *Blaise Pascal* to analyze *gambling* games (Ben-Naim, 2008).

For example (Ben-Naim, 2008), there was a popular die game where the players would each choose a number from 1 to 6. The die would be thrown consecutively until the number chosen by one player appeared 3 times. The game would end and this player would be the winner. The problem to be solved was how to split the money in the pot if the game had to be stopped before there was a winner. For example if a player's number had appeared once and the number of another had appeared twice, and then the police stopped the game, how should they split the money?

Axiomatic probability theory

In his 1933 monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung* (in English *Foundations of the Theory of Probability*, 1956) Soviet mathematician *Andrey Nikolaevich Kolmogorov* presented the axiomatic basis of modern probability theory. According to Koutsoyiannis (1997) and Ben-Naim (2008) probability theory is constructed by three basic concepts and three axioms.

The basic concepts are:

Sample space

The *sample space* Ω is a set whose elements ω are all the possible outcomes of an *experiment* (or *trial*).

For example, for the throw of a die the sample space is:

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Likewise the sample space for the throw of two *distinguishable* dice is:

Ω = {[1 1], [1 2], [1 3], [1 4], [1 5], [1 6], [2 1], [2 2], [2 3], [2 4], [2 5], [2 6], [3 1], [3 2], [3 3], [3 4], [3 5], [3 6], [4 1], [4 2], [4 3], [4 4], [4 5], [4 6], [5 1], [5 2], [5 3], [5 4], [5 5], [5 6], [6 1], [6 2], [6 3], [6 4], [6 5], [6 6]}.

Note that the outcomes {[1 2]} and {[2 1]} are different due to the distinguishability of the two dice.

Events

The subsets of Ω are called *events*. An event A occurs when the outcome ω of the experiment is an element of A. The sample space is also called the *certain event*. The empty set Ø is also called the *impossible event*.

From the first example above, the event “less than or equal to 2” is

$$A = \{1, 2\}.$$

From the second example above, the event “the sum is equal to 7” is

A = {[1 6], [2 5], [3 4], [4 3], [5 2], [6 1]}.

The event “the sum is equal to 1” is

$$A = \emptyset.$$

The *field of events* $\mathcal{F}$ is a set of all the subsets A of Ω, including[5] Ω itself and the empty set.

Measure of probability

The *measure of probability* P is a real function on $\mathcal{F}$. This function assigns to every event A a real number $P(A)$ called *the probability of A*.

The three elements $\{\Omega, \mathcal{F}, P\}$ together define a *probability space*.

The three axioms

The following three axioms define the properties that the measure of probability P must fulfill.

1. $$P(\Omega) = 1 \tag{2.1}$$

2. $$0 \leq P(A) \leq 1 \tag{2.2}$$

These two axioms define the range of the function $P(A)$, and the fact that the certain event has the largest probability measure.

3a. $$P(A \cup B) = P(A) + P(B), \text{ if } A \cap B = \emptyset \tag{2.3}$$

3b. $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)\text{ , if } A_i \cap A_j = \emptyset, \text{ for } i \neq j. \tag{2.4}$$

According to Koutsoyiannis (1997) axiom 3b is introduced as a separate axiom because it is not a consequence but an extension of 3a to infinity. The events A and B of 3a (or A_i of 3b) are said to be *disjoint* or *mutually exclusive*.

Two immediate consequences of the three axioms are presented below.

Probability of the impossible event

Given that

$$\Omega \cup \emptyset = \Omega \quad \text{and} \quad \Omega \cap \emptyset = \emptyset$$

and given axioms 1 and 3a,

$$P(\emptyset) = 0 \tag{2.5}$$

[5] According to set theory a set is always subset of itself; the empty set is a subset of all sets.

Non-disjoint events

For two non-disjoint events A and B it is shown that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.6}$$

Two important concepts of probability theory are the *independence of events* and *conditional probabilities*.

Independent events

According to Ben-Naim (2008), two events are said to be independent if the occurrence of one event has no effect on the probability of occurrence of the other. In mathematical terms, two events are called independent if and only if

$$P(A \cap B) = P(A)\, P(B) \tag{2.7}$$

For n independent events we have

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)\, P(A_2) \dots P(A_n) \tag{2.8}$$

Conditional probabilities

The conditional probability of an event is the probability of its occurrence *given* that another event has occurred.

The conditional probability is denoted and defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{2.9}$$

If the two events are independent then from 2.7 and 2.9:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\, P(B)}{P(B)} = P(A) \tag{2.10}$$

which means that the occurrence of B has no effect on the occurrence of A, which is what we would expect for independent events.

Definitions of probability

Even though 80 years have passed since Kolmogorov presented the three simple axioms on which the whole mathematically theory of probability is based, the *meaning* of the concept of probability remains an open problem for mathematicians and philosophers. Here I present some clarifications regarding the *definition* of probability.

According to Ben-Naim (2008),

> *In the axiomatic structure of the theory of probability, the probabilities are said to be* assigned *to each event. These probabilities must subscribe to the three*

> *conditions a, b and c* [axioms 1, 2, 3a and 3b of this thesis]. *The theory does not* define *probability, nor provide a method of calculating or measuring these probabilities. In fact, there is no way of calculating probabilities for any general event. It is still a quantity that measures our degree or extent of belief of the occurrence of certain events. As such, it is a highly subjective quantity.*

The two most common definitions of probability are presented by Ben-Naim (2008). These definitions provide ways to calculate probabilities and are called the *classical definition*, and the *relative frequency definition*.

The classical definition

The classical definition is also called *a priori* definition. A priori is Latin for "from the earlier", in the sense of "from before", "in advance". According to this definition probabilities are calculated before any measurements of the experiment at hand are made.

Following Ben-Naim's notation, if $N(total)$ is the total number of outcomes of an experiment (i.e. the number of elements of the field of events $\mathcal{F}$) and $N(event)$ the number of outcomes (i.e. the number of elementary events) composing the event of interest, then the probability of the event is calculated by the formula

$$P(event) = \frac{N(event)}{N(total)} \tag{2.11}$$

This definition is based on deductive reasoning: starting from what is known about the experiment (numbers of events) we reach a conclusion about the probabilities of possible events. It is based on the *principle of indifference* of Laplace, according to which two events are to be assigned equal probabilities if there is no reason to think otherwise (Jaynes, 1957). More specifically this definition is based on the assumption that all elementary events have the same probability. This assumption makes the classical definition circular (Ben-Naim, 2008) since it assumes that which it is supposed to calculate. For example, as stated earlier, the probability of the outcome "4" when we throw a die is 1/6. But why do we believe that each of the six outcomes of a fair die should have the same probability of occurrence?

The classical definition, having the form of formula 2.11, is only one specific way to calculate a priori probabilities. In subsection 2.6, a different much more powerful and rigorous way to calculate a priori probabilities will be presented based on the *principle of maximum entropy*. The classical definition will be shown to be just a special case of this principle and we will be able to answer why each outcome of a die throw has probability 1/6.

The relative frequency definition

This definition is also called the *a posteriori* or *experimental* definition. A posteriori is Latin for "from the later", in the sense of "in hindsight", "in retrospect". According to this definition probabilities are calculated based on measurements, i.e. *after* the experiment.

Following Ben-Naim's (2008) example, we consider the toss of a coin. We denote the two possible events as H (head) and T (tail). We toss the coin N times and count how many time H occurs, denoting the number of occurrences as $n(H)$. The frequency of occurrence of H is equal to $n(H)/N$. According to the relative frequency definition the probability of H is the limit of the frequency of occurrence when the total number of tosses N tends to infinity, i.e.

$$P(H) = \lim_{N \to \infty} \frac{n(H)}{N} \tag{2.12}$$

This definition is not problem-free either. First, it is not possible to repeat an experiment an infinite number of times, and second, even if it were possible, there is no guarantee that the limit of formula 2.12 will converge. Therefore this definition is used for a very large N, without having the certainty however, of how large is large enough.

Random variables and probability distributions

A *random variable* is a function X defined on a sample space Ω that associates each elementary event ω with a number $X(\omega)$ according to some *predefined rule*. The outcome ω may be a number and the predefined rule a mathematical function (Koutsoyiannis, 1997). For example ω may be the reflectivity measured by a rainfall radar and the predefined rule may be the formula transforming the reflectivity into rainfall intensity. But the outcome and the rule may be much more abstract. For example ω may be the color of the t-shirt of the first person that Aris sees every morning and the rule may be assigning the values "1" for "*red*", "2" for "*blue*", "3" for "*green*" and "4" for "*all other colors*".

Usually we omit the element ω and simply write X, unless for clarity reasons we cannot omit it. To denote the random variable *itself* we use capital letters while to denote a *value* of the random variable we use small letters. For example we write $\{X \leq x\}$ meaning the event that is composed of all elementary events ω such that the values $X(\omega)$ are less than or equal to the number x. The probability of such an event is denoted $P(\{X(\omega) \leq x\})$ or more simply $P(X \leq x)$ (Koutsoyiannis, 1997).

Distribution functions

According to Koutsoyiannis (1997), the *distribution function* F_X is a function of x defined by the equation

$$F_X(x) = P(X \leq x),\ x \in \mathbf{R},\ F_X \in [0,1] \tag{2.13}$$

F_X is not a function of the random variable X, it is a function of the real number x. Thus X is used as an index of F, i.e. we write F_X, not $F(X)$. We use the index to differentiate between distributions of various random variables. If there is no danger of confusion we can omit the index.

Furthermore, the *domain* of F is not identical to the *range* of $X(\omega)$ but is always the whole set of real numbers **R**. F is always an increasing function, following the inequality

$$0 = F_X(-\infty) \leq F_X(x) \leq F_X(+\infty) = 1 \tag{2.14}$$

F_X is also called *cumulative distribution function* (CDF) or *non-exceedance probability*.

If $F_X(x)$ is continuous for all x, then the random variable $X(\omega)$ is called *continuous*. In this case the sample space Ω is an infinite and uncountable set. On the other hand if $F_X(x)$ is a step function, then the random variable $X(\omega)$ is called *discrete*. In this case the sample space Ω is a finite set or an infinite and countable set. It is important to note however that even for discrete random variables, the CDF is always defined for all $x \in \mathbf{R}$.

For continuous random variables, the derivative of the CDF is called *probability density function* (pdf) and is given by

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{2.15}$$

The cumulative distribution function can be calculated by the inverse of equation 2.15, which is

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)d\xi \tag{2.16}$$

The variable ξ is just a real number. We use it within the integral so that we can use x as the integration limit.

The main properties of the probability density function are

$$f_X(x) \geq 0 \tag{2.17}$$

and

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{2.18}$$

Appendix A presents additional properties of random variables, namely distribution functions (marginal, joint and conditional), expected values, and moments.

Stochastic processes[6]

The theory of *stochastic processes* is an application of probability theory. It can be used to describe the temporal evolution or the spatial relations of random variables. The formal definition of a stochastic process is that it is "*a family of random variables*" (Koutsoyiannis, 1997).

A stochastic process is denoted as X_t where the index t takes values from an appropriate index set T. T can refer to time, for example the stochastic process can be the temperature of an area. But T can also refer to space or any other set. For example the stochastic process can be the concentration of a pollutant along the length of a river.

Furthermore T's dimension can be greater than one. In this case the stochastic process is usually called a *random field*. In this case if $\mathbf{v} = \{v_1, v_2, \ldots, v_n\}$ is a vector of the n-dimensional index set $\mathbf{V}$, then a random field is defined as the family of random variables $X(\mathbf{v})$. The random variable may be scalar (i.e. be of the form $\mathbf{R}^n \rightarrow \mathbf{R}$) or have more dimensions (i.e. be of the form $\mathbf{R}^n \rightarrow \mathbf{R}^m$). An example of a scalar 2-dimensional random field is a landscape, where the vector $\mathbf{v}$ is the coordinates of a point and the random variable is the altitude. An example of a scalar 3-dimensional random field is the air temperature and an example of a vectorial 3-dimensional random field is the windspeed.

The greatest advantage of using stochastic processes for the study of random phenomena instead of just using statistics is that we can take into account the temporal or spatial *dependence* between random variables.

For example if we are interested in successive throws of a die we do not need the theory of stochastic processes given that each throw is independent. But if we are interested in the

[6] This subsection is based on Koutsoyiannis (1997) and Theodoratos (2004)

average daily discharge of a large river we can use stochastic processes. We can reasonably expect that the discharge of a day will statistically depend on the discharge of the previous day. We can use certain stochastic processes to model this dependence.

Distribution functions

The distribution function of the random variable X_t is defined as

$$F(x;t) = P(X(t) \leq x) \tag{2.19}$$

and is also called *1st order distribution function* of the stochastic process.

We can also define higher order distribution functions as joint distribution functions of n random variables X_t. For example the function

$$F(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = P(X(t_1) \leq x_1 \cap X(t_2) \leq x_2 \cap \ldots \cap X(t_n) \leq x_n) \tag{2.20}$$

is called the n^{th} *order distribution function* of the stochastic process.

The *mean*, or *expected value*, of a stochastic process is defined as

$$\mu(t) = \mathrm{E}[\, X(t)\,] = \int_{-\infty}^{\infty} x\, f(x;t)\, \mathrm{d}x \tag{2.21}$$

The second order joint covariance is called *autocovariance* and is given by

$$\mathrm{Cov}[\, X(t_1), X(t_2)\,] = \mathrm{E}[\, (X(t_1) - \mu(t_1))(\, X(t_2) - \mu(t_2))\,] \tag{2.22}$$

Stationarity

A stochastic process is *stationary* when the probability distributions of all orders of the random variable X_t are constant for all t. Mathematically this can be expressed as:

$$f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = f(x_1, x_2, \ldots, x_n; t_1 + c, t_2 + c, \ldots, t_n + c), \text{ for all } n, c \tag{2.23}$$

A stochastic process is called *wide-sense stationary* when the probability distributions of X_t are not necessarily constant, but the mean is constant and the autocovariance depends only on the difference $\tau = (t_1 - t_2)$. Mathematically this can be expressed as:

$$\mathrm{E}\, [\, X(t)\,] = \mu = \text{constant} \tag{2.24}$$

and

$$\mathrm{Cov}[\, X(t), X(t+\tau)\,] = \mathrm{E}[\, (X(t) - \mu)(\, X(t+\tau) - \mu)\,] = C(\tau) \tag{2.25}$$

Ergodicity

A stochastic process is called *ergodic* when its statistical properties (e.g. its mean) can be estimated from a single sample of the stochastic process of sufficient length. Mathematically, for a discrete process this can be expressed as:

$$\mathrm{E}[\, X(t)\,] = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N} X(t) \tag{2.26}$$

and for a continuous process as:

$$\mathrm{E}[\, X(t)\,] = \lim_{T\to\infty} \frac{1}{T} \int_{0}^{T} X(t)\mathrm{d}t \tag{2.27}$$

Ergodicity is a very important property. Only when a stochastic process is ergodic we can perform measurements to infer its statistical properties. If it is not ergodic then no matter how long a measurement record is we cannot calculate its statistics.

2.2 Information theory: Shannon's Entropy

Information theory is a branch of applied mathematics. It was established by the American mathematician and electrical engineer *Claude Elwood Shannon* while he was working at Bell Laboratories, and presented in his 1948 paper titled *A Mathematical Theory of Communication*. Information theory was developed to study the sending and receiving of messages on telecommunication lines. After its introduction it found many applications in the field of electrical and computer engineering, helping engineers solve problems regarding the storage, compression and transmission of signals and data. However concepts of information theory are now used in very diverse fields of science, from biology to linguistics and from economics to ecology.

The *Shannon entropy*, a measure of information, is one of the most basic concepts of information theory. Before giving the formal definition of Shannon entropy I discuss the concept of information.

Information

The word *information* is used widely in our daily lives. It has various meanings for various persons in various circumstances. It can be used colloquially or formally, subjectively or objectively.

In the context of information theory when we use the expression “the information carried by a message” we are not interested in the *content* of the message but on the message itself, specifically on the *size* of the message and the *probability* of the message being what it is. The content of the message may be important or trivial; it may be true or false; it may be useful or useless; it may even be completely meaningless, just a random series of letters. This is irrelevant for information theory. Only the message itself is relevant to information theory.

According to Shannon

> *The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have* meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one* selected from a set *of possible messages.*

Telecommunications and stochastic processes

Information theory could be seen as a branch of stochastic processes. It uses concepts such as probabilities, random variables, Markov processes[7], etc. This may be surprising at first. After all, telecommunications are based on devices and systems designed, controlled and used by humans in very specific ways to transmit signals. Stochastic processes on the other hand are dealing with processes that occur randomly. It is very useful however to view a communication signal as a random process. The reason is that even though the sender of a signal knows what was sent, the receiver does not. Therefore from the point of view of the receiver, the signal is a *random signal*. In addition no system is perfect, therefore the signal is transmitted with *noise*, i.e. with random, small or large, perturbations. Therefore when designing telecommunication systems we treat signals as random. Telecommunications and stochastic processes nowadays are so closely linked that many times even natural time series, such as rainfall time series, are called signals.

---

[7] A Markov process is a stochastic process with the property that each successive random variable depends only on its previous random variable and not on random variables further back. An important consequence of this property is that if we know the present value of a Markov process we can predict its future with the same uncertainty as if we knew its entire past. Mathematically this can be described by the equation

$$X(i) = aX(i\text{-}1) + V(i) \qquad (2.42)$$

where $X(i)$ the random variable at time i, a a constant that gives rise to the dependence of $X(i)$ to $X(i\text{-}1)$, and $V(i)$ white noise, i.e. an independent random variable. (Koutsoyiannis, 2007)

Shannon's work on information theory was based on earlier contributions of *Harry Nyquist* and *Ralph Hartley*. But Shannon formalized the existing theory, expanded it to include noisy transmissions, introduced stochastic approaches, and, most importantly, discovered the information measure of entropy.

In the beginning of his paper, Shannon introduces the basic definitions of information theory, such as what is a signal, an information source, a channel, noise, etc.

Then he presents the applicability of stochastic processes within information theory. Using an example he explains various ways that a transmitted written message can be treated by a probabilistic approach. He assumes that the "alphabet" used for the messages has five letters. First he assumes that each successive selection of a letter is independent and all five letters are equiprobable. Then he assumes that letters have different probabilities. Making the example more complicated he assumes that successive symbols are not independent, but have some correlation to one or more previous selections. Finally he expands the example to include the random creation and selection of "words". A second example expands the previous to the 27 symbols of the English alphabet (26 letters and the space). The processes that he described in these examples are discrete Markov processes.

Thus the ground had been laid for the introduction of the concept of entropy. But before presenting the concept of entropy I would like to make two interesting side-notes regarding the first sections of Shannon's paper.

First, the first time the term "bit" was used in engineering was by Shannon in this paper. He gives credit to J. W. Tukey for suggesting it. The term bit comes from the combination of the words "binary digit". A binary digit is a unit of information and results from using a logarithm with base 2 as a measure of information. The bit has become the typical unit of information due to the fact that digital computers are based on the so-called "flip-flop circuits", which are devices that can be in two states. Therefore, as Shannon explains, N such devices can store N bits, since the total number of possible states is 2^N and $\log_2 2^N = N$.

Second, 100 years before Shannon's paper was published, *Edgar Allan Poe* presented a methodology that can be used to decipher an English text written with numbers and symbols instead of letters. His methodology was based on a zero-order Markov process, i.e. he used only the relative frequencies of English letters, and some intuitive reasoning to determine the dependence structure of characters and words. It appeared in 1843 in the short-story "*The Gold-Bug*". Perhaps Poe should be given some credit for the development of information theory. In fact, according to some sources, for example an interview he gave in 1982 to Robert Price[8], Shannon's interest in the field was influenced by reading the Gold-Bug as a boy.

Shannon's Entropy

The fourth section of Shannon's paper was titled "Choice, uncertainty and entropy". He wanted to quantify the uncertainty involved in the transmission of a message. Shannon first defined what he wanted to quantify. He asked:

> *Can we define a quantity which will measure, in some sense, how much information is "produced" by such a Markoff process, or better, at what rate information is produced?*
>
> *Suppose we have a set of possible events whose probabilities of occurrence are p_1, $p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?*

Basic properties of the uncertainty measure

Then Shannon defined three properties that we can reasonably demand this measure to have. He wrote:

> *If there is such a measure, say $H(p_1, p_2, \ldots, p_n)$, it is reasonable to require of it the following properties:*
>
> 1. *H should be continuous in the p_i.*

We can reasonably expect that for small changes of the probabilities p_i the change of the measure *H* should also be small.

> 2. *If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.*

It is intuitively clear, as Shannon explains, that as *n* increases there is more uncertainty as to which event occur. Therefore the measure of uncertainty is expected to increase.

> 3. *If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Fig. 6* [Figure 2.1 of this thesis]. *At the left we have three possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. On the right we first choose between two*

[8] The interview can be read in *http://www.ieeeghn.org/wiki/index.php/Oral-History:Claude_E._Shannon*

possibilities each with probability $\frac{1}{2}$*, and if the second occurs make another choice with probabilities* $\frac{2}{3}, \frac{1}{3}$*. The final results have the same probabilities as before. We require, in this special case that*

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right)$$

The coefficient $\frac{1}{2}$ *is because this second choice only occurs half the time.*

This requirement is equivalent to demanding that the measure of uncertainty depends only on the distribution of p_i and not on our way of finding out which event occurs. In other words it is a requirement for objectivity. It clarifies that H measures the uncertainty that is inherent in the experiment *itself* and not *our* uncertainty.

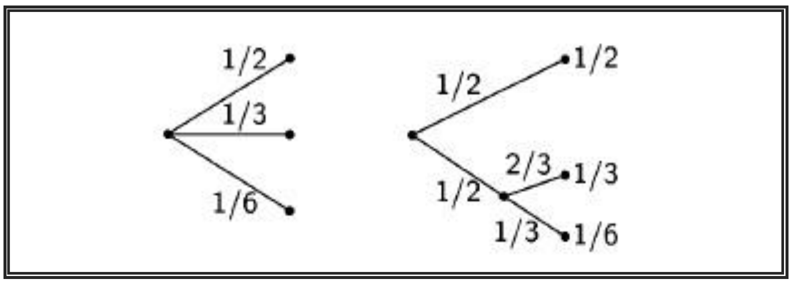


Figure 2.1: Decomposition of a choice from three possibilities. Source: Shannon (1948)

Formula of the measure of uncertainty

Shannon then proved[9] that the only function that satisfies these properties is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \tag{2.28}$$

K is any positive constant. The base of the logarithm can be any number. In many cases the choice of some specific logarithmic base can greatly simplify calculations. In information theory frequently the base is chosen to be equal to 2 and the unit of the resulting measure of uncertainty is a bit. In statistical physics the base is usually the number e, i.e. the logarithm is the natural logarithm (ln).

[9] Shannon's original proof can be found in Appendix 2 of Shannon (1948). An additional proof can be found in Appendix F of Ben-Naim (2008).

We can easily see by equation 2.28 that the measure of uncertainty is the expected value of the function

$$g(p_i) = -K\log p_i \tag{2.29}$$

therefore

$$H = \mathrm{E}[-K\log p_i\,] \tag{2.30}$$

Even after proving equation 2.28 Shannon did not realize that his measure of uncertainty had the same form as the thermodynamic entropy of Gibbs. Before publishing his paper however he had the discussion with Von Neumann that was mentioned in the Introduction. Thus he wrote:

> *The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell i of its phase space. H is then, for example, the H in Boltzmann's famous H theorem. We shall call* $H=-\sum_{i=1}^{n} p_i \log p_i$ *the entropy of the set of probabilities $p_1,\ldots, p_n$.*

Finally, regarding notation Shannon clarifies that:

> *If x is a chance [i.e. random] variable we will write H(x) for its entropy; thus x is not an argument of a function but a label for a number, to differentiate it from H(y) say, the entropy of the chance variable y.*

Shannon used the letter H as the symbol of entropy to honor Boltzmann and his famous H-theorem[10].

Shannon's entropy can be seen as a function that manages to summarize an entire probability distribution with a single number. More specifically, the quantity $\log p_i$ can be seen as a quantity that measures how likely is the occurrence of the event i, the only difference with p_i is that instead of mapping the likelihood of i in the range [0,1], it maps it in the range (-∞,0]. Based on this, the quantity $\sum p_i \log p_i$ can be seen simply as the *average likelihood* of all the events i. Therefore its opposite, i.e. Shannon's entropy, is nothing more than the *average unlikelihood* of all events i. Therefore, when comparing two different distributions, it is natural that the uncertainty will be higher for the distribution with a high average unlikelihood.

[10] Boltzmann's H-theorem is presented in the next section.

The discovery of Shannon's entropy

The way that Shannon discovered the measure of uncertainty is very interesting from an epistemological point of view. Shannon was searching for a measure of the size of information as technological developments in telecommunications were pushing for a formalization of the theory of information. He used common sense to come up with three very simple properties that the measure must satisfy. Then he discovered that there is only one function that satisfies these properties, so he adopted it. Furthermore he found that this function has a few more properties, which according to him

> *further substantiate it as a reasonable measure of choice or information.*

These and other important properties are presented in the following section.

Notwithstanding the debate regarding the relation – or non-relation – between the information-theory and thermodynamic entropy, it should be clear that there is nothing "magic" or metaphysical in the form of equation 2.28. It is nothing more than what the mathematics gave to Shannon when he asked for three simple properties, based on common-sense.

Properties of Shannon entropy[11]

The certain event

It can be easily proved that for any distribution p_i, $H = 0$ if and only if one p_i is equal to one (certain event) and all other p_i are equal to zero.

For any distribution p_i, $i = 1, 2, \ldots, n$, we know that,

$$0 \leq p_i \leq 1$$

and

$$-\infty < \log p_i \leq 0$$

It follows that if $p_i \neq 1$, or in other words if $0 \leq p_i < 1$ for all p_i, then

$$\log p_i < 0 \;\Rightarrow\; -\log p_i > 0 \;\Rightarrow\; -p_i \log p_i \geq 0 \;\Rightarrow\; H = -\sum_{i=1}^{n} p_i \log p_i > 0^*$$

[11] The formulation of the properties is after Shannon (1948), Ben-Naim (2008). The notation may be different.

* The inequality $-p_i \log p_i \geq 0$ becomes $H > 0$ because not all p_i can be equal to zero, so their sum cannot be equal to zero.

While if $p_j = 1$ for a certain j and $p_i = 0$ for all other i, then

$$\log p_j = 0$$

therefore

$$p_i \log p_i = 0 \text{ for all } i, \text{ including } j$$

and

$$H = -\sum_{i=1}^{n} p_i \log p_i = 0$$

This property is a consequence of the fact that log1 = 0. It is absolutely reasonable to expect that a reasonable measure of uncertainty should be zero when the outcome is certain!

Equiprobable events (uniform distribution)

For a given distribution p_i, $i = 1, 2, \ldots, n$, uncertainty H is maximum when all the events i are equiprobable, and have probabilities

$$p_i = \frac{1}{n} \tag{2.31}$$

This is proved using the method of *Lagrange multipliers*[12].

This is also an intuitively expected property. It is reasonable to expect that the most uncertain experiment is one where all outcomes are equally probable. If some of the events are more probable, the uncertainty for the outcome is less.

The uncertainty corresponding to the uniform distribution is given by

$$H_{\max} = -\sum_{i=1}^{n} p_i \log p_i = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = -\log \frac{1}{n} = \log n \tag{2.32}$$

Joint entropy

Let X and Y be two random variables with distributions

$$p_i = P_X(i) = P(x_i) = P\{X = x_i\},\ i = 1, 2, \ldots, n$$

and

$$q_j = P_Y(j) = P(y_j) = P\{Y = y_j\},\ j = 1, 2, \ldots, m$$

Also let the joint probability distribution be

$$P_{ij} = P_{XY}(i \cap j) = P_{XY}(\mathrm{x}_i \cap \mathrm{y}_j) = P\{(X = x_i) \cap (Y = y_j)\}$$

[12] The proof is presented in Appendix B

The joint entropy, i.e. the entropy of the compound event is[13]

$$H(X \cap Y) = -\sum_{i,j} P_{ij} \log P_{ij} \tag{2.33}$$

The marginal distributions of X and Y are

$$p_i = \sum_{j=1}^{m} P_{ij} \tag{2.34}$$

and

$$q_j = \sum_{i=1}^{n} P_{ij} \tag{2.35}$$

Their entropies are

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i = -\sum_{i,j} P_{ij} \log \sum_{j=1}^{m} P_{ij} \tag{2.36}$$

and

$$H(Y) = -\sum_{j=1}^{m} q_j \log q_j = -\sum_{i,j} P_{ij} \log \sum_{i=1}^{n} P_{ij} \tag{2.37}$$

It can be easily shown that

$$H(X \cap Y) \leq H(X) + H(Y) \tag{2.38}$$

The equality holds when the two random variables are independent, i.e. when

$$P_{ij} = p_i q_j \tag{2.39}$$

The meaning of this property is that the joint uncertainty for the combined outcome of two experiments cannot be larger than the sum of uncertainties for each experiment alone, which also follows common sense.

Conditional entropy

For dependent random variables a conditional measure can be defined.

[13] Sometimes the joint entropy is written as $H(X, Y)$

Let X and Y be two dependent random variables. The conditional probability of the event y_j given that the event x_i has occurred is given by

$$P(y_j \mid x_i) = \frac{P(x_i \cap y_j)}{P(x_i)} \qquad (2.40)$$

The conditional entropy of Y given X is defined as the average entropy of Y for each value of X, weighted according to the probability $P(x_i)$ of each x_i., or, in other words, it is defined as the expected value of $H(Y \mid X)$:

$$H(Y \mid X)\ ] = \sum_i P(x_i)\mathrm{H}(Y \mid x_i) = \sum_i P(x_i)\left(-\sum_j P(y_i \mid x_i)\log P(y_i \mid x_i)\right) =$$

$$= -\sum_{i,j} P(x_i)\, P(y_i \mid x_i)\log P(y_i \mid x_i) = -\sum_{i,j} P(x_i)\frac{P(x_i \cap y_j)}{P(x_i)}\log\frac{P(x_i \cap y_j)}{P(x_i)} =$$

$$= -\sum_{i,j} P(x_i \cap y_j)\log\frac{P(x_i \cap y_j)}{P(x_i)} = -\sum_{i,j} (\, P(x_i \cap y_j)\log P(x_i \cap y_j) - P(x_i \cap y_j)\log P(x_i)\,) =$$

$$= -\sum_{i,j} P(x_i \cap y_j)\log P(x_i \cap y_j) + \sum_{i,j} P(x_i \cap y_j)\log P(x_i) = H(X \cap Y) + \sum_i \left(\sum_j P(x_i \cap y_j)\right)\log P(x_i)) =$$

$$= H(X \cap Y) + \sum_i P(x_i)\log P(x_i) = H(X \cap Y) - H(X)$$

We have shown that

$$H(Y \mid X) = H(X \cap Y) - H(X) \qquad (2.41)$$

Combining this equation with inequality 2.38 we get

$$H(X \cap Y) \le H(X) + H(Y) \Leftrightarrow H(Y \mid X) + H(X) \le H(X) + H(Y) \Leftrightarrow H(Y \mid X) \le H(Y) \quad (2.42)$$

Infinite outcomes

When n tends to infinity the uncertainty function may or may not be defined.

For example for equiprobable events,

$$\lim_{n\to\infty} H = \lim_{n\to\infty} \log n = \infty \qquad (2.43)$$

even though

$$\lim_{n\to\infty} p_i = \lim_{n\to\infty} \frac{1}{n} = 0 \qquad (2.44)$$

For non-equiprobable events the definability of the uncertainty function depends on whether the limit

$$\lim_{n\to\infty}\left(- \sum_{i=1}^{n} p_i \log p_i \right) \tag{2.45}$$

converges or diverges.

Continuous variables

For a variable with continuous distribution $f(X)$ the uncertainty can be defined as

$$H(X) = \int_{-\infty}^{\infty} f(x)\log f(x)\mathrm{d}x \tag{2.46}$$

i.e.

$$H(X) = \mathrm{E}[\ \log f(x)\ ] \tag{2.47}$$

According to Ben-Naim (2008) the uncertainty function of continuous variables may have problems of non-convergence and different properties than the discrete case. Examples are discussed in Appendix I of his book.

Havrda – Charvat – Tsallis entropy

According to Papalexiou and Koutsoyiannis (2012), Shannon's entropy was generalized by Havrda and Charvat in 1967 and reintroduced by Tsallis in 1988.
Havrda – Charvat – Tsallis entropy is defined as:

$$H_q^{\mathrm{HaChTs}}(p_1,\ldots,p_n) = \frac{1 - \sum_{i=1}^{n} p_i^{\,q}}{q - 1} \tag{2.48}$$

It is easy to show that

$$\lim_{q\to 1} H_q^{\mathrm{HaChTs}} = H^{\mathrm{Sh}} \tag{2.49}$$

where H^{Sh} is Shannon's entropy.

For a non-negative continuous random variable X, Havrda – Charvat – Tsallis entropy is defined as:

$$H_q^{\mathrm{HaChTs}}(X) = \frac{1 - \int_{0}^{\infty} (\ f(x)\ )^{q}}{q - 1} \tag{2.50}$$

2.3 Thermodynamics: Clausius', Boltzmann's and Gibbs' Entropy

Thermodynamics is the branch of physics that studies heat, its relation to other forms of energy and the relation of both to the properties of matter (Moran and Shapiro, 2000; Borel and Favrat, 2010). A word of Greek origin, thermodynamics means *dynamics of heat*. It originates in the early 19th century and was developed to study the conversion of heat into work by steam engines. But steam engines were only the trigger. As the *laws of thermodynamics*, especially the first and second, were formulated and their ramifications explored, it became clear that the field of application of thermodynamics is much wider and covers all natural sciences. It studies the most diverse processes, from the heart of the stars to the nuclei of living cells. *Entropy* and the *second law of thermodynamics*, in the formulation of which the concept of entropy first appeared, attempt to explain *how* and *why* changes in the Universe take place.

Clausius' Entropy

The term *entropy* was coined by German physicist and mathematician *Rudolf Clausius* in 1863 (Ben-Naim, 2008) in the context of classical thermodynamics. He introduced the concept of entropy to answer some very pressing engineering problems of his time.

Conservation of energy

The *first law of thermodynamics* guarantees the conservation energy. Energy cannot be destroyed or created, it can only be transformed. For example if an engine is using some amount of energy, E_{used}, to produce work, W, while some energy, E_{lost}, is lost due to friction, heat leaks, etc., then the following equality must hold

$$E_{\text{used}} = E_{\text{lost}} + W \tag{2.51}$$

The *efficiency* of such an engine is defined as the ratio of the work produced over the energy used by the engine

$$\alpha = \frac{W}{E_{\text{used}}} \tag{2.52}$$

It follows that

$$\alpha = \frac{W}{E_{\text{used}}} = \frac{E_{\text{used}} - E_{\text{lost}}}{E_{\text{used}}} = 1 - \frac{E_{\text{lost}}}{E_{\text{used}}} \tag{2.53}$$

Entropy and the second law of thermodynamics were developed as scientists and engineers were trying to find ways to increase the efficiency α of steam engines, or equivalently to reduce the amount of lost energy E_{lost}.

Carnot's efficiency formula

French engineer *Sadi Carnot*, who is considered to have laid the foundations of thermodynamics, studied the problem of engines' efficiency. He tried to imagine an *ideal heat engine* that can operate without any friction or heat leaks. A heat engine is an abstraction of steam engines. It consists of a *hot source* of heat, such as a boiler, a device that transforms heat into work, such as a piston, and a *cold sink* where excess heat is dumped, such as a cold water tank.

At the time other engineers were experimenting with various steam engine designs, using various substances instead of steam or water (Atkins, 2010). Carnot managed to prove that the efficiency of an ideal heat engine, or in other words the maximum possible efficiency of any real engine, depends *only* on the *temperatures* of the hot source and the cold sink, according to the formula

$$\alpha = 1 - \frac{T_{\mathrm{sink}}}{T_{\mathrm{source}}} \tag{2.54}$$

where T is the absolute temperature.

Carnot's formula has a very important implication. Even the most perfect engine cannot have 100% efficiency, as this would require either T_{sink} to be zero or T_{source} to be infinite, both of which are impossible. But his work had very little impact during his time as it was contrasting the prevailing engineering views of the time (Atkins, 2010).

The second law of thermodynamics

The second law of thermodynamics was first formulated by Clausius in 1849, while a second formulation was presented in 1850 by British physicist and engineer *William Thomson* – later known as *Lord Kelvin* (Ben-Naim, 2008). Even though their formulations were different, they can be shown to be equivalent.

Clausius was studying the flow of heat from hot to cold bodies (Atkins, 2010). This is an everyday phenomenon, but Clausius realized that it signifies an asymmetry in nature. While the flow of heat from a hot to a cold body is *spontaneous*, the *reverse* process is not spontaneous. His statement of the second law was

> *heat does not pass from a body at low temperature to one at high temperature without an accompanying change elsewhere.*

In other words heat can flow from a cold to a hot body but only if it is facilitated by the expenditure of work. This is the "accompanying change elsewhere" that Clausius refers to.

Kelvin was studying heat engines (Atkins, 2010). He realized that the cold sink is essential for a heat engine to operate. Before him the cold sink was largely ignored, as engineers were mostly focused on the hot source or the piston. In fact some engineers did not even realize that their designs had a cold sink, in cases for example that steam was released in the air and the environment played the role of the sink. Kelvin's statement of the second law however explained that

> *no cyclic process is possible in which heat is taken from a hot source and converted completely into work.*

In other words part of the heat that is taken from the source must be deposited to a sink. This statement has the same implication as Carnot's formula; no heat engine can have 100% efficiency.

Atkins (2010) shows that the two statements are equivalent. He proves that if we assume that heat can flow spontaneously from cold to warm – that is, if Clausius' statement is false – then a cyclic process that converts all heat to work is possible – that is, Kelvin's statement is also false. Then he proves the reverse, if Kelvin's statement is false, then Clausius' statement must also be false.

Entropy

Entropy was introduced by Clausius to give a *mathematical form* to the second law of thermodynamics. Most probably when he introduced it he could not imagine how scientists' understanding of entropy would expand and evolve. His idea was simply that he could define a *new state variable* that would allow the law to be defined by a simple inequality instead of the two statements formulated by him and Kelvin. He used various formulations and names between 1854 and 1865 for entropy. The formulas presented today by most textbooks are not necessarily identical to his original ones.

For the purposes of this thesis a simple formulation of Clausius' entropy can be the following.

If there is flow of a *small* amount of heat ΔQ in a system at absolute temperature T[14] then the *change of entropy* of the system, ΔS, is given by the equation

$$\Delta S = \frac{\Delta Q}{T} \tag{2.55}$$

We adopt the convention that if heat is *absorbed* then ΔQ is *positive*, while if heat is *released* then ΔQ is *negative*. Given that absolute temperature is always positive, the sign of ΔS is the same as the sign of ΔQ.

[14] The amount of heat is assumed small so that the temperature can be considered constant

The law can be reformulated using the change of entropy. For example:
The total entropy of a system and its surroundings increases during any spontaneous process.

Or in mathematical form

$$\sum \Delta S \geq 0 \tag{2.56}$$

The equality holds for ideal *reversible* processes. Real processes are *irreversible*, for which the inequality holds.

To show that this formulation is equivalent to Clausius' original statement we assume an isolated system composed of two sub-systems, a hot source with absolute temperature T_1 and a cold sink with absolute temperature T_2. Let ΔQ be an amount of heat that flows spontaneously from the source to the sink, in accordance with the second law of thermodynamics. The changes of entropy corresponding to the source and sink are

$$\Delta S_1 = \frac{-\Delta Q}{T_1} \quad \text{and} \quad \Delta S_2 = \frac{\Delta Q}{T_2} \tag{2.57}$$

In accordance with the sign convention, the flow of heat out of the source has a negative sign. Therefore the entropy of the source is reduced. The opposite is true for the sink.
We have the following inequalities

$$T_2 < T_1 \;\Leftrightarrow\; \frac{1}{T_2} > \frac{1}{T_1} \;\Leftrightarrow\; \frac{\Delta Q}{T_2} > \frac{\Delta Q}{T_1} \;\Leftrightarrow\; \Delta S_2 > -\Delta S_1 \;\Leftrightarrow\; \Delta S_2 + \Delta S_1 > 0 \;\Leftrightarrow\; \sum \Delta S > 0$$

What I find interesting with Clausius' definition of entropy is that due to the properties of numbers, namely due to the fact that if $T_2 < T_1$ then $\frac{1}{T_2} > \frac{1}{T_1}$, a very simple formula manages to capture such an important law of nature. The amount of heat flowing out of the hot source causes a decrease of its entropy. But when the same amount flows to the cold sink it causes an increase of its entropy that is larger. Thus there is a net increase in entropy for the total process.

Atkins (2010) compares this very illustratively with a whisper in a busy street versus a whisper in a library. The busy street is like the source, the whisper does not make much of a difference; the library is like the sink, the same whisper makes a big difference.

To show that the formulation using entropy is equivalent to Kelvin's statement we imagine a cyclic process that violates the second law and converts all heat into work. The entropy of the hot source will be reduced by $\Delta S_1 = \frac{-\Delta Q}{T_1}$, as described above. But there will be no increase of entropy elsewhere in the system or its surroundings, since changes of entropy are associated with heat, not with work.

This example shows why the cold sink is so essential. Without the sink it would not be possible for the system to compensate for the decrease of the source's entropy. Using inequality 2.56 we can calculate the minimum heat that must be absorbed by the sink.

Let Q_1 be the heat extracted from the source (which has temperature T_1) and Q_2 the heat absorbed by the sink (which has temperature T_2). The respective changes of entropy are

$$\Delta S_1 = \frac{-Q_1}{T_1} \text{ and } \Delta S_2 = \frac{Q_2}{T_2} \tag{2.58}$$

We have the following inequalities

$$\sum \Delta S \geq 0 \iff \Delta S_2 + \Delta S_1 \geq 0 \iff \Delta S_2 \geq -\Delta S_1 \iff \frac{Q_2}{T_2} \geq \frac{Q_1}{T_1} \iff Q_2 \geq Q_1 \frac{T_2}{T_1}$$

Or equivalently

$$Q_{2\,\mathrm{min}} = Q_1 \frac{T_2}{T_1} \tag{2.59}$$

Based on the above we can calculate the maximum work that can be performed by an ideal heat engine and its maximum possible efficiency.
Let W be the work, given by

$$W = Q_1 - Q_2 \tag{2.60}$$

We have the following inequalities

$$Q_2 \geq Q_1 \frac{T_2}{T_1} \iff -Q_2 \leq -Q_1 \frac{T_2}{T_1} \iff Q_1 - Q_2 \leq Q_1 - Q_1 \frac{T_2}{T_1} \iff W \leq Q_1 \left(1 - \frac{T_2}{T_1}\right)$$

Or equivalently

$$W_{\mathrm{max}} = Q_1 \left(1 - \frac{T_2}{T_1}\right) \tag{2.61}$$

Dividing both sides by Q_1 we get

$$\alpha_{\mathrm{max}} = 1 - \frac{T_2}{T_1} \tag{2.62}$$

which can be recognized as Carnot's efficiency formula.

Implications of the second law of thermodynamics

The second law of thermodynamics has some important implications.

Maximum entropy at equilibrium

The first implication is that,

the entropy of an isolated system that is in equilibrium is constant and has reached its maximum value.

This statement is sometimes used as a formulation of the law. It is easy to see why it holds using inequality 2.56.

To prove the statement we first assume that the system is at a state out of equilibrium. As it moves towards equilibrium its entropy will increase, therefore at the original state its entropy was not at a maximum.

On the other hand if we assume that the system has maximum entropy then its entropy cannot increase further, which means that no spontaneous process can take place. If no spontaneous process can take place it means that the system is at equilibrium.

Dissipation of energy

Every process results in the dissipation of heat to the environment. In the ideal case, for example in the case of Carnot's ideal heat engine, the amount of heat released to the environment (the cold sink) is equal to the minimum heat given by equation 2.59. In a real case the heat released would be even more due to friction, and other losses. This heat cannot be used for work. Therefore, somehow misleadingly, it is usually referred to as "low-quality energy". Thus, while the *quantity* of energy remains constant, its *quality* is always reduced.

Heat death of the Universe

If we consider the entire Universe as a thermodynamic system, the dissipation of energy, taken to its extreme, will lead to the transformation of all energy to "low-quality" heat which will be unable to perform any work. The Universe will become a uniform reservoir of heat. German physicist *Hermann von Helmholtz* named this the "heat death" of the Universe (Ball, 2004).

The arrow of time

The second law gives rise to the arrow of time, a term coined by British astrophysicist *Arthur Stanley Eddington* to describe the fact that time has a direction.

In classical mechanics all processes are reversible. We can think for example an elastic ball with mass *m* held at a height *h* above the floor. It has zero kinetic energy and potential energy equal to *mgh*. If it is released it will fall and when it reaches the floor it will have kinetic energy equal to *mgh*, i.e. velocity equal to $\sqrt{2gh}$, and zero potential energy. Then it will bounce on the floor and will return to height *h* with zero velocity, zero kinetic energy, and potential energy equal to the initial *mgh*. The process is reversed. If we watch a movie of this process there is no way of telling if the movie is played in reverse or not.

In thermodynamics however, even for an ideal Carnot heat engine, a certain amount of energy, given by equation 2.59, is always lost. In thermodynamics, heat always flows from a hot to a cold body. This *direction* of processes creates the arrow of time. If we see a movie of a glass of water in room temperature that starts to "spontaneously" warm up we will know that we are watching the movie in reverse.

We cannot use a refrigerator to cool a room
Finally, if there is a local decrease of entropy it must be compensated with an increase of entropy somewhere else. For real – i.e. not ideal – processes the increase will *overcompensate* the decrease, i.e. the total entropy ('local' plus 'somewhere else') will increase. This shows why we cannot cool a room by leaving the door of a refrigerator open!

The atomic theory of matter and Boltzmann's Entropy

When Carnot developed his thermodynamic theories, the atomic theory of matter was not yet the prevailing view. He considered that "*heat was a kind of imponderable fluid that, as it flowed from hot to cold, was able to do work, just as water flowing down a gradient can turn a water mill*" (Atkins, 2010). Thermodynamic quantities and functions, such as heat, temperature, pressure, and entropy, referred to solid bodies and bulk quantities of liquids and gasses. This kind of thermodynamics is called *classical*. For classical thermodynamics it is irrelevant whether matter is made up of atoms. But as the atomic theory and, later on, quantum mechanics were developed, classical thermodynamics gave way to *statistical mechanics*. This subsection presents entropy from the point of view of statistical mechanics. The discussion however is mostly qualitative as any exact derivation would exceed both my knowledge and the scope of this thesis.

The atomic theory of matter
The *atomic theory of matter* has a very long history. In the 4th and 5th centuries BCE, Ancient Greek philosophers *Leucippus*, most famously, *Democritus* and *Epicurus*, believed that all matter was composed of atoms. Much later, in 1738 *Daniel Bernoulli* published

Hydrodynamica where he argued that gases consist of molecules, whose motion and impact on surfaces cause what we perceive as heat and pressure. The theory was proved, and became finally accepted, by the combined efforts of *Albert Einstein* and *Jean Perrin*. Einstein provided a theoretical explanation of *Brownian motion*[15] in 1905 and Perrin confirmed experimentally Einstein's theory in 1908 (Ball, 2004).

The kinetic theory of gases

Even before the atomic theory was widely accepted, great minds of their time developed the *kinetic theory of gases*. The groundwork for the theory was laid to a large degree by Clausius and was further developed by Scottish physicist *James Clerk Maxwell*. Maxwell's key contribution was the idea that instead of a detailed description of the trajectory and momentum of each and every particle of a gas we could just use equations of the average behavior of particles, more specifically of the average and the variance of the particles' velocity. He assumed that for a gas *in equilibrium* the velocities depend only on the temperature of the gas and their probability distribution follows a bell-shaped curve. This distribution is called *Maxwell-Boltzmann distribution* and can be seen in figure 2.2. From the figure we can see that as the temperature increases so do both the average and the variance of the velocity. (Ball, 2004)

[15] Brownian motion is the random motion of tiny particles, visible with strong microscopes. It was first discovered *Robert Brown* in 1828 (Ball, 2004).

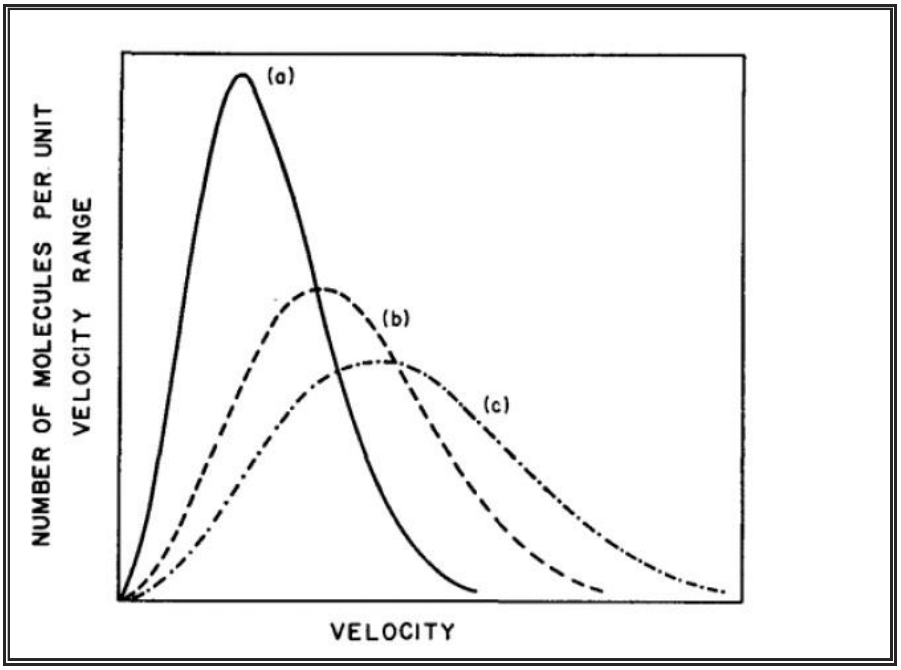


Figure 2.2: Maxwell-Boltzmann distribution of molecule velocities of a gas at temperatures *a*, *b*, *c*, with *c* > *b* > *a*. Source: Dugdale (1996)

Boltzmann's equations

Maxwell's derivations were intuitive. The distribution was confirmed in a rigorous mathematic manner by another great mind, the Austrian physicist *Ludwig Eduard Boltzmann*. Boltzmann derived the distribution of energy of the particles. The distribution of energy amongst particles is not continuous because each particle can have only certain energies as specified by quantum mechanics. There are discrete levels of energy and each level has a number of particles that occupy it. If N is the total number of particles, N_i is the number of particles with energy level E_i, and T is the absolute temperature in Kelvin degrees, then the Boltzmann energy distribution is

$$\frac{N_i}{N} = e^{-E_i/kT} \tag{2.63}$$

Where k is the Boltzmann constant, equal to 1.38×10^{-23} J/K.

From the distribution of energies Boltzmann showed that the distribution of velocities is the same that Maxwell predicted.

Furthermore, Boltzmann added a very important element to the theory. Maxwell's derivation was defined for a gas in equilibrium. Boltzmann on the other hand showed that for any initial distribution of velocities, when the gas reaches equilibrium the velocities will follow the Maxwell-Boltzmann distribution. This introduced the concept of change in the kinetic theory. Boltzmann realized that the distribution of velocities was related to entropy, the variable that describes change in thermodynamic systems.

Thus Boltzmann derived a new formula for entropy at equilibrium. First he made the distinction between the *macrostate* and the *microstate* of the system. The macrostate is the state that can be described macroscopically by variables such as heat, temperature, pressure, etc. The microstate is the configuration of the locations and velocities of all the particles of the system. A macrostate can arise from a very large number of microstates. This means that even though the microstates are constantly changing, as vast numbers of particles collide with each other, change position, and exchange momentum, at a macroscopic level the system seems to be in equilibrium. Boltzmann denoted the number of possible microstates by W. Furthermore he assumed that at equilibrium all microstates are equiprobable. Based on these assumptions he derived the equation (Dugdale, 1996):

$$S = k \ln W \tag{2.64}$$

where k is the Boltzmann constant.

The assumption that microstates are equiprobable follows from the so-called *ergodic hypothesis* of Boltzmann. In simple words the ergodic hypothesis states that a particle may be at all available locations with equal probability and may have all possible velocities with equal probability.

Given that microstates are equiprobable, it is easy to see that the probability of the system to be in a microstate i is

$$p_i = \frac{1}{W} \tag{2.65}$$

From this it follows that the Boltzmann entropy equation can be rewritten as

$$S = k \ln \frac{1}{p_i} = - k \ln p_i \tag{2.66}$$

New perspectives on entropy

Boltzmann's equation resulted in new perspectives on and interpretations of the second law of thermodynamics and the concept of entropy.

First of all it gave a *statistical notion* to entropy. Macroscopic properties of a system, such as heat, temperature, and pressure, were already shown to be statistical properties that arise by the behavior and interactions of a huge number of particles. Now entropy was shown to also be such a property.

Equation 2.64 shows that the entropy of a system is an increasing function of W. This means that entropy's increase during spontaneous processes leads the system to macrostates that can be achieved in more ways, or in other words that arise from more microstates. Consequently equilibrium can be defined as the macrostate that arises from the *maximum number of microstates*, given that the entropy is maximum at equilibrium.

For example figure 2.3 shows the expansion of an isolated gas. In the left, the gas is at equilibrium and confined by a partition to the left half of a container. When the partition is removed the gas will eventually reach a new equilibrium state where it will occupy the whole container. We can intuitively understand that since there is more volume available for each particle, there are more microstates that give rise to the new equilibrium macrostate, therefore the entropy of the gas will increase. As a consequence of the Sackur-Tetrode equation, this entropy increase is given by:

$$\Delta S = N \ln \frac{V}{0.5V} = N \ln 2 > 0 \tag{2.67}$$

where V is the total volume of the box and N is the number of particles of the gas (Ben-Naim, 2008).

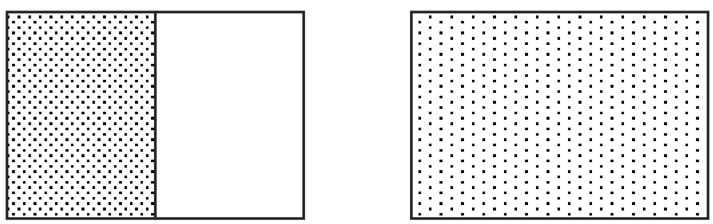


Figure 2.3: Expansion of a gas from an initial volume 0.5 *V* to a final volume *V*.

Since during spontaneous processes the entropy cannot decrease, a gas that occupies a container will not *spontaneously confine* itself to one part of the container. If the process is not spontaneous the entropy of the system can be reduced by the expenditure of work, i.e. by increasing entropy somewhere else. For example we can use a piston to *compress* the gas back to the left half of the container.

Examples like this have contributed to the view that entropy is a measure of "*disorder*". The reasoning behind this view is that when all particles of a gas are gathered in one side of their container then they are more ordered, while when they fill the whole container, they are spread-out, they are less ordered

On the other hand, looking at equation 2.66, we can easily see how entropy can be associated with *uncertainty*. When there is more uncertainty about the actual microstate that the system is in, i.e. when the probability P_i is less, then entropy is increased.

Maxwell's demon

Boltzmann's breakthroughs presented very serious philosophical challenges to scientists of the late 19th century. According to the prevailing philosophical and religious beliefs, the universe was harmonious, symmetrical and beautiful. The laws of nature were given by God, while humans were purposeful and had free will. These views were seriously challenged by the second law of thermodynamics and especially by Boltzmann's statistical definition of entropy.

For example Maxwell, a deeply religious man (Ball, 2004), tried to prove through a thought experiment that the second law was a human illusion, and that an intelligent being with free-will could be free from this law. This being was named *Maxwell's Demon* by Kelvin, perhaps as an irony towards Maxwell's religiousness. Ben-Naim (2008) provides the following quote of Maxwell, according to which,

> *Starting with a uniform temperature, let us suppose that such a vessel is divided into two portions or by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole so as to allow only the swifter molecules to pass from A to B, and only the slower ones pass from B to A. He will thus, without expenditure of work raise the temperature of B and lower that of A in contradiction to the second law of thermodynamics.*

It has been proven however, both from a quantum-mechanics and an information-theoretic point of view, that such a being would need to spend energy or create entropy to be able to measure the particles' velocities. Therefore the total entropy of the gas plus the demon would still be increased.

Irreversibility arising from reversibility: Boltzmann's H-theorem

Another very troubling question was how is it possible to have irreversible macroscopic processes arising from reversible microscopic dynamics. Let us think for example the case of the expanding gas of figure 2.3. When the partition is removed, it is the collisions of particles to each other and the walls that make them expand to fill the whole container. But the dynamics of these collisions are reversible. Why is then the expansion of the gas irreversible?

This line of thought is what philosophers call *mechanistic*. It tries to reduce a more complex phenomenon to simple mechanics. Reductions are abstractions, designed to simplify a complexity. They are simplified imaginary examples of a real phenomenon. While reductions are many times helpful in illuminating something complex, when we make them we cannot expect the original phenomenon to behave like the reduced phenomenon. We must not forget that our abstractions are not real and expect reality to follow our simplified example.

The mechanistic view was prevalent at Boltzmann's time. So his theory was met with widespread skepticism. In response to this skepticism Boltzmann introduced the *H-theorem* to show how reversible dynamics of gas particles can lead to irreversible macroscopic processes. The H-theorem is described below in very simplified terms based on Brown et al. (2009).

Let $f(\mathbf{r}, \mathbf{v}, t)$ be a density function, where $\mathbf{r}$ is the position vector, $\mathbf{v}$ is the velocity vector and t is time. The product $f(\mathbf{r}, \mathbf{v}, t)\mathrm{d}^3\mathbf{r}\mathrm{d}^3\mathbf{v}$ gives the number of particles that are contained within an infinitesimal volume element $\mathrm{d}^3\mathbf{r}$ centered around $\mathbf{r}$ and have velocities contained within the element $\mathrm{d}^3\mathbf{v}$ and centered around $\mathbf{v}$. Due to the very large number of particles f is assumed to be a continuous function.

Boltzmann first defined a *transport equation* that describes how $f(\mathbf{r}, \mathbf{v}, t)$ varies with time under the influence of external forces, of diffusion and of forces between particles due to collisions. In the derivation of the H-theorem Boltzmann focused only on forces due to collisions.

Furthermore – and very importantly – he assumed that particles that are about to collide are *uncorrelated*. Based on this assumption the joint density function of a pair of particles that are about to collide is given by

$$F(v_1, v_{2,}\, \mathrm{t}) = f(v_1, t)\, f(v_2, t) \tag{2.68}$$

Then he defined the H function, given by

$$H\left[\, f_t \,\right] = \int f(\mathbf{r}, \mathbf{v}, t) \ln f(\mathbf{r}, \mathbf{v}, t)\, \mathrm{d}^3\mathbf{r}\mathrm{d}^3\mathbf{v} \tag{2.69}$$

Boltzmann showed, first that H can only decrease, i.e.

$$\frac{\mathrm{d}H}{\mathrm{d}t} \leq 0 \tag{2.70}$$

and second that for N particles, entropy S can be defined in terms of the H function by the equation

$$S = -\, N\, k\, H \tag{2.71}$$

where k is the Boltzmann constant.

Therefore entropy can only increase. This way Boltzmann proved mathematically how irreversibility arises out of reversibility.

Despite the *H*-theorem the mechanistic and religious views of his contemporaries were very difficult to shatter. Boltzmann hanged himself on September 5^{th} 1906. It is believed that amongst the reasons that led to his suicide was his disillusionment by the prevalence of positivism amongst scientists. The formula

$$S = k \cdot \log W$$

is inscribed on his tombstone.

Even though Boltzmann views on the composition of matter have been confirmed, his *H*-theorem has been criticized on various grounds. For example it has been criticized for ignoring external forces and diffusion, i.e. for assuming that the gas is already uniformly distributed. Also he has been criticized for assuming that the particles are uncorrelated.

Why is there a Second Law of Thermodynamics?

The second law of thermodynamics is considered as one of the most fundamental laws of nature. However the question "why is there such a law" is still an open question as there is a number of possible explanations. One of these explanations, which I find convincing and easy to grasp, lies in its statistical nature. According to this explanation, it is overwhelmingly more probable for entropy to increase. But there is nothing in the laws of nature that forbids it to be reduced. If we are willing to wait long enough we would see all particles of the gas of figure 2.3 return to the left half of the container. But we would need to wait so long that the age of the Universe would probably not be long enough.

Ben-Naim (2008) explains that

> *the system will spend more time at events (or states) of higher probability. This, in itself, provides an explanation of the "driving force" for the evolution of the system; from a state of low to a state of high probability.* […] *In this view, we have not only an interpretation for the quantity S as MI* [missing information; equivalent to uncertainty as explained in Chapter 1], *but also an answer to the question of "why" the system evolves in the specific direction of increasing the entropy (or the MI). In this view, the statement of the Second Law is reduced to a statement of common sense, nothing more.*

He further explains that,

> *one should realize that the Second Law is crucially dependent on the atomic constituency of matter. If matter were not made up of a huge number of atoms, the Second Law would not have existed, and the entropy would not have been defined. The same is true for the concepts of temperature and heat. Hence, also the Zeroth and the First Law of Thermodynamics would not have been formulated.*

Gibbs' Entropy

American physicist *Josiah Willard Gibbs* refined Boltzmann's definition of entropy. He extended the definition to include non-equilibrium macrostates. The microstates that correspond to non-equilibrium macrostates are not necessarily equiprobable. Therefore instead of using the logarithm lnp_i as Boltzmann did (equation 2.66), Gibbs uses the *expected value* of lnp_i. Therefore his definition of entropy is

$$S = -k \sum_i p_i \ln p_i \qquad (2.72)$$

Statistical ensembles

In addition to their different formulations, Gibbs' and Boltzmann's entropies have a deeper conceptual difference. Boltzmann's definition refers to microstates that the system can actually be in. Gibbs' definition on the other hand refers to *statistical ensembles*. A statistical ensemble is a fictitious collection of copies of the system. It is not a state that the system is really in; it is all the states that theoretically the system could be in.

The two points of view are identical if the ergodic hypothesis holds. But if it does not hold they may be different. The consequences of this difference play an important role on the discussion of irreversibility and other theoretical questions of statistical mechanics, such as the *fluctuation theorem* (Callender, 1999), which expresses analytically the probability of a system *violating* the second law (Evans and Searles, 2002).

The macroscopic constraints on the system determine the properties and probability distributions of the microscopic ensembles. The most important ensembles are the *microcanonical ensemble*, which refers to isolated systems (i.e. systems that cannot exchange energy nor matter with their environment) at constant temperature; the *canonical ensemble*, which refers to closed systems (i.e. systems that can exchange energy but cannot exchange matter with their environment) at constant temperature; and the *grand-canonical ensemble*, which refers to open systems at constant temperature.

Other entropies

There are many other entropy expressions defined by various scientists to capture various phenomena.

One example is the Von Neumann entropy. It was defined by Hungarian mathematician *John Von Neumann* to extend the concept of entropy to the field of quantum mechanics.

Entropy production

So far I have been mostly considering equilibrium thermodynamics. A thermodynamic system is said to be in equilibrium when, from a macroscopic point of view, its properties, such as pressure, temperature, concentrations, etc., are *uniform* across space and remain *constant* in time. In such a system there are *no gradients*.

However most systems in nature are not in equilibrium. These systems try to reach equilibrium by *dissipating the gradients* that cause them to be out of equilibrium. As they do so they *produce entropy* in accordance with the second law.

There are cases however that the production of entropy takes place in the surroundings of the system and the local entropy in the system is reduced, such as living systems, creating the illusion that the law is violated. Such cases are presented in Chapter 3.

Non-equilibrium thermodynamics studies systems out of equilibrium. One of the pioneers of the field was Norwegian chemist and physicist *Lars Onsager*. He was particularly interested in steady-state systems (Ball, 2004). These systems, while they are not in equilibrium, still maintain some properties constant in time.

Non-equilibrium thermodynamics also studies phenomena that give rise to forms and patterns, such as phase transitions (e.g. the formation of snowflakes), or dissipative structures (e.g. turbulence, or Benard cells)

A central concept of non-equilibrium thermodynamics is the *entropy production rate*, sometimes called just entropy production.
It is defined simply as

$$\frac{\mathrm{d}S}{\mathrm{d}t} \tag{2.73}$$

Physicists have been searching for general laws that can describe non-equilibrium systems and the *paths* that they take as they move towards equilibrium. These laws would be able for example to explain and predict steady states, forms and patterns. Onsager won the

Nobel Prize of Chemistry in 1968 for his work on systems that are very close to equilibrium. But not much progress has been made towards finding an all-encompassing principle for non-equilibrium systems, which, as many believe, would be related to the entropy production rate. According to Ball (2004) "*there is a good reason for this: it is almost certain that no such principle exists*."

Russian-Belgian chemist *Ilya Romanovich Prigogine* believed that he found such a principle. He showed that processes in some systems near equilibrium occur in such a way that the entropy production is *minimum*, or in other words, these systems approach equilibrium at the *slowest* possible rate. According to Ball (2004) this criterion "*is not universally true*".

Maximum entropy production (rate)

Another principle is the *maximum entropy production (rate) principle* proposed by *Ziegler*. This principle at first appears to be in contradiction with Prigogine's principle. But as Kleidon et al. (2010) explain, Prigogine's principle refers to transient (i.e. not steady), near-equilibrium processes, while Ziegler's principle refers to steady-state, far-from-equilibrium processes. Other sources, such as Martyushev and Seleznev (2006), agree that the two principles are not in contradiction.

Ziegler's principle states that a system will choose the path that maximizes entropy production as it moves towards equilibrium, or in other words it will approach equilibrium at the *fastest* possible rate.

The implications of the extremization of entropy production are very important and far reaching. Some of them will be presented in Chapter 3. But the use and recognition of the maximum entropy production principle has been delayed due to the fact that scientists researching this topic were largely unaware of each other's work as Martyushev and Seleznev (2006) wrote in a literature-review paper which, according to them, was the first review paper on the topic ever to be published.

The maximum entropy production principle has not been proven. In fact according to Martyushev (2010) "*a principle like MEPP cannot be proved*." But this does not necessarily mean that this principle is less important. Martyushev (2010) writes:

> *Examples of its successful applications for description of observed phenomena just support this principle, while experimental results (if they appear) contradicting the principle will just point to the region of its actual applicability. The balance of the positive and negative experience will eventually lead to the consensus of*

> *opinion on the true versatility or a limited nature of MEPP. Other principles, such as laws of thermodynamics and Newton's law, developed along similar lines.*

Sometimes this is how science progresses.

2.4 The Principle of Maximum (Statistical) Entropy

The *Principle of Maximum Entropy* was introduced by American physicist *Edwin Thompson Jaynes* (1957). It is a method of *statistical inference* based on Shannon's entropy. The motivation for the introduction of the principle was to derive statistical mechanics distributions and equations based solely on rules of statistical inference and the basic laws of physics (e.g. conservation of energy) without the need to introduce additional assumptions, such as ergodicity. More specifically the method was introduced so that "*information theory can be applied to the problem of justification of statistical mechanics*" (Jaynes, 1957).

The introduction of information entropy by Shannon and the fact that he showed it having

> *a deeper meaning* […] *independent of thermodynamics* […] *makes possible a reversal of the usual line of reasoning in statistical mechanics. Previously, one constructed a theory based on the equations of motion, supplemented by additional hypotheses of ergodicity, metric transitivity, or equal a priori probabilities, and the identification of entropy was made only at the end, by comparison of the resulting equations with the laws of phenomenological thermodynamics. Now, however, we can take entropy as our starting concept, and the fact that a probability distribution maximizes the entropy subject to certain constraints becomes the essential fact which justifies use of that distribution for inference.* (*Ibid.*)

Since its introduction, the maximum entropy principle has found application in many fields, including hydrology. Based on this principle we can explain for example why the probability to get a certain outcome when throwing a fair die is 1/6. More examples of applications are presented in Chapter 3.

According to Jaynes (*Ibid.*) the maximum-entropy estimate is "*the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information*". This means that it uses all of the available information (e.g. measurements) but without using any information that is not available (e.g. by making assumptions regarding the distribution that a random variable should follow).

Or in other words (*Ibid.*),

> *In the problem of prediction, the maximization of entropy is not an application of a law of physics, but merely a method of reasoning which ensures that no unconscious arbitrary assumptions have been introduced.*

Description of the methodology

A general description of the inference problem and the methodology to solve it is presented below based on Jaynes (*Ibid.*). The notation used here differs at times.

Let X be a random variable taking discrete values x_i, i= 1, 2, …, n. The distribution of probabilities p_i is not given. All that is known is the expected value of a function $g(X)$

$$\mathrm{E}[g(X)] = \sum_{i=1}^{n} p_i g(x_i) \tag{2.74}$$

Based on this information, what is the expected value of another function y(X)?

Jaynes writes (*Ibid.*):

> *At first glance, the problem seems insoluble because the given information is insufficient to determine the probabilities p_i. Equation* [2.74] *and the normalizing condition*
>
> $$\sum_{[i=1]}^{[n]} p_i = 1 \qquad [(2.75)]$$
>
> *would have to be supplemented by (n-2) more conditions before* [$E[\, y(X)\,]$] *could be found.*

However information theory provides a solution to this problem. Shannon's entropy of X can be used. As we saw earlier it is given by

$$H(p_1, p_2, \ldots, p_n) = - \sum_{i=1}^{n} p_i \ln p_i \tag{2.76}$$

Jaynes considered the development of information theory 9 years earlier by electrical engineer Claude Shannon a very important opportunity for physics. He noted that this importance was felt by others, but it was not clear how to apply the new theory of information to the field of statistical physics.

Jaynes explained that,

> *The great advance provided by information theory lies in the discovery that there is a unique, unambiguous criterion for the "amount of uncertainty" represented by a discrete probability distribution.*

And he continues:

> *It is now evident how to solve our problem; in making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have.*

Therefore the problem becomes a *maximization problem* and can be solved easily by employing the method of *Lagrange multipliers* (Efstratiadis and Makropoulos, 2011).

We are looking for that distribution of p_i that maximizes H of equation 2.76, subject to the constraints of equations 2.74 and 2.75.

We define the auxiliary function

$$\varphi(p_i, \lambda_0, \lambda_1) = H(p_1, p_2, \ldots, p_n) - \lambda_0 \sum p_i - \lambda_1 \sum_{i=1}^{n} p_i g(x_i) \tag{2.77}$$

We search for p_i such that

$$\frac{\partial \varphi(p_i, \lambda_0, \lambda_1)}{\partial p_i} = 0 \tag{2.78}$$

which yields

$$\ln p_i = -1 - \lambda_0 - \lambda_1 g(x_i) \tag{2.79}$$

Given that λ_0 is a constant we can replace $1+\lambda_0$ with a new λ_0 to simplify the formulas.

Thus H is maximum for

$$p_i = \mathrm{e}^{- \lambda_0 - \lambda_1 g(x_i)} \tag{2.80}$$

i.e.

$$H_{\max} = - \sum_{i=1}^{n} p_i \ln p_i = - \sum_{i=1}^{n} p_i \ln \mathrm{e}^{- \lambda_0 - \lambda_1 g(x_i)} = - \sum_{i=1}^{n} p_i (- \lambda_0 - \lambda_1 g(x_i)) = \lambda_0 \sum_{i=1}^{n} p_i + \lambda_1 \sum_{i=1}^{n} p_i g(x_i)$$

which, according to equations 2.74 and 2.75, becomes

$$H_{\max} = \lambda_0 + \lambda_1 \mathrm{E}[\, g(x_i) \,] \tag{2.81}$$

Knowing the distribution of p_i we can now calculate the expected value of $y(X)$ by the equation

$$\mathrm{E}[\, y(X)\,] = \sum_{i=1}^{n} p_i\, y(x_i) \tag{2.82}$$

If we have more information, we can refine the distribution given by equation 2.80.

For example for the general case of knowing the expected values of m functions g_m the distribution of p_i becomes

$$p_i = \mathrm{e}^{-\lambda_0 - \lambda_1 g_1(x_i) - \ldots - \lambda_\mathrm{m} g_m(x_i)} \tag{2.83}$$

And the H maximum is

$$H_{\max} = \lambda_0 + \lambda_1 \mathrm{E}[\, g_1(x_i)\,] + \ldots + \lambda_m \mathrm{E}[\, g_m(x_i)\,] \tag{2.84}$$

Example: Uniform distribution

As already discussed in the section of the properties of Shannon's entropy, when our only information is the number of possible outcomes, i.e., when the only constraint is equation 2.75, then entropy is maximized for a uniform distribution.

An outcome of fair die throw has probability 1/6 because the uniform distribution maximizes the entropy.

Example: Gaussian distribution

Koutsoyiannis (2005a) gives the following example.

We assume that we know the mean, μ, and standard deviation, σ, of a continuous random variable X, or using the notation used above,

$$\mathrm{E}[\, g_1(X)\,] = \mathrm{E}[\, X\,] = \mu \tag{2.85}$$

and

$$\mathrm{E}[\, g_2(X)\,] = \mathrm{E}[\, X^2\,] = \mu_2 \tag{2.86}$$

The quantity μ_2 is the *order 2 raw moment*[16] and is related with the mean and standard deviation according to the relation

$$\mu_2 = \mu^2 + \sigma^2 \tag{2.87}$$

[16] See Appendix A for the definition of moments.

The entropy of X is

$$H(X) = \int_{-\infty}^{\infty} f(x) \ln f(x) \mathrm{d}x \tag{2.88}$$

Also, we have the constraint

$$\int_{-\infty}^{\infty} f(x) \mathrm{d}x = 1 \tag{2.89}$$

Application of the principle of maximum entropy with the three constraints given by 2.85, 2.86, and 2.89, yields:

$$H_{\max} = \lambda_0 + \lambda_1 \mu + \lambda_2 \mu_2 \tag{2.90}$$

and

$$f(x) = \mathrm{e}^{-\lambda_0 - \lambda_1 x - \lambda_2 x^2} \tag{2.91}$$

Algebraic manipulations show that the Lagrange multipliers and the maximum entropy are given by

$$\lambda_0 = \ln\left(\sigma\sqrt{2\pi}\right) + \frac{\mu^2}{2\sigma^2} \tag{2.92}$$

$$\lambda_1 = -\frac{\mu}{\sigma^2} \tag{2.93}$$

$$\lambda_2 = \frac{1}{2\sigma^2} \tag{2.94}$$

$$H_{\max} = \ln\left(\sigma\sqrt{2\pi\mathrm{e}}\right) \tag{2.95}$$

and that equation 2.91 is the *Gaussian* (or *normal*) distribution.

Koutsoyiannis (2005a) notes that it is interesting that the entropy of a random variable following the normal distribution depends only on its standard deviation and not on its mean.

Philosophical views of Jaynes

Jaynes' 1957 paper has many interesting implications about statistical inference and logic; nature; and probability theory. A number of extensive quotes of this paper are presented in Appendix C of this thesis.

2.5 Thermodynamic vs. Information Entropy

The discovery of the uncertainty function by Shannon and the naming of it as entropy, has led to a debate as to whether his entropy is the same concept as thermodynamic entropy or not. This is still an open debate. A quick search of the literature yields many publications on the topic.

One view is that the two entropies are not to be confused with each other as they are unrelated concepts. I believe that it is caused, amongst other reasons, by the following two.

The first reason is historic. It has to do with the fact that thermodynamic entropy has units of J/K, while Shannon's entropy is dimensionless (Koutsoyiannis, 2011b). It is very difficult therefore for many scientists to comprehend how a dimensionless quantity can be identical to a quantity with units.

But thermodynamic entropy does not need to have units. As Atkins explains (Koutsoyiannis, 2011b),

> *although Boltzmann's constant k is commonly listed as a fundamental constant, it is actually only a recovery from a historical mistake. If Ludwig Boltzmann had done his work before Fahrenheit and Celsius had done theirs, then … we might have become used to expressing temperatures in the units of inverse joules… Thus, Boltzmann's constant is nothing but a conversion factor between a well-established conventional scale and the one that, with hindsight, society might have adopted.*

If we were using 1/J as the unit of temperature then Boltzmann's constant would be dimensionless, and so would be thermodynamic entropy.

It is interesting to note however that Atkins is amongst those who do not believe that the two entropies are the same concept. This brings us to the second reason for this view. Ben-Naim (2008) quotes the following paragraph from Atkins.

> *I have deliberately omitted reference to the relation between information theory and entropy. There is the danger, it seems to me, of giving the impression that entropy requires the existence of some cognizant entity capable of possessing 'information' or of being to some degree 'ignorant.' It is then only a small step to the presumption that entropy is all in the mind, and hence is an aspect of the observer. I have no time for this kind of muddleheadedness and intend to keep such metaphysical accretions at bay. For this reason I omit any discussion of the analogies between information theory and thermodynamics.*

Similarly, Wicken (1987) compares the use of entropy in thermodynamics and information theory. He writes at one point,

> *when noise is introduced, increases in message uncertainty correlate with increases in both its entropy and its information content, since greater "freedom of choice" among messages is made available […] While the maximally noisy channel delivers maximum possible information, it is also useless information – that is, information that conveys no information.*

At another point he writes

> *The paradox here, noted by many over the years, is that if entropy is a state property of a system it cannot depend on what we happen to know about that system.*

It seems that Atkins and Wicken view information as something subjective, in the sense that it depends on someone's knowledge. According to this view, different observers with different knowledge would calculate different amounts of information. However the word "information" in the context of information theory does not refer to the content of a message but to the series of characters that compose the message itself, as it was explained in Chapter 2 of this thesis. Therefore "information" has nothing to do with our knowledge – or anyone else's knowledge – or with the content of messages or systems.

In physics there is another kind of subjectivity which is not related to observers but to reference frames. For example Koutsoyiannis (2011a) explains that the location coordinates of points are subjective, as they depend on the coordinate frame. But the distance between points is objective, as it does not depend on the coordinate frame. Koutsoyiannis (*Ibid.*) proves that there can be a probability-based definition of thermodynamic entropy that does not depend on the frame of reference. Therefore this definition of entropy is objective in the reference-frame sense.

Interestingly Jaynes, who used Shannon's entropy to derive thermodynamic entropy, was amongst those who believed that thermodynamic and information-theory entropy were different concepts and should not be confused. Koutsoyiannis (2011b) quotes him saying that,

> *We must warn at the outset that the major occupational disease of this field is a persistent failure to distinguish between the information entropy, which is a property of any probability distribution, and the experimental entropy of thermodynamics, which is instead a property of a thermodynamic state as defined, for example by such observed quantities as pressure, volume, temperature, magnetization, of some physical system. They should never have been called by the*

> *same name; the experimental entropy makes no reference to any probability distribution, and the information entropy makes no reference to thermodynamics. Many textbooks and research papers are flawed fatally by the author's failure to distinguish between these entirely different things, and in consequence proving nonsense theorems*

There is also the opposite view in this debate, namely that the two concepts are related. Both are a measure of uncertainty; in the one case uncertainty of a telecommunications transmission, or some other information technology application, while in the other case uncertainty of the microstates of a thermodynamic system.

Ben-Naim (2008), who supports this view, compares entropy with the cosine function. He explains that

> *the fact that cosθ appears in two different fields, say in the propagation of an electromagnetic wave, and in the swinging of a pendulum, does not imply a deep and profound connection between the two fields. However, in both fields, the appearance of cosθ indicates that the phenomena are periodic.*

Similarly, he explains, the fact that the function $-\sum p_i \log p_i$ appears in two different fields does not mean that there is a deep connection between the fields. But it does mean that in both fields this function measures uncertainty.

Koutsoyiannis (2011a) shows how the laws of thermodynamics can be derived by using only the concept of statistical entropy and the principle of the maximization thereof, with the addition of conservation laws and Newton's second law to derive constraints. In effect what Koutsoyiannis shows is that not only are the two entropies closely related, but thermodynamic entropy is in fact a special case of the more general concept of statistical entropy.

I find more convincing the view that the two concepts are closely related. Furthermore I believe that using Shannon's entropy it is much easier to see that entropy is a measure of uncertainty. This can then offer a much clearer insight on what thermodynamic entropy is, making it more difficult to misunderstand it as a measure of "disorder".

3. Applications of Entropy

In this chapter I present several applications where the concept of entropy, in its various definitions presented in the previous chapter, is used. I choose applications where the concept of entropy plays a major role. I also present some consequences of the properties of entropy that will help create a better understanding of the concept and its huge importance for nature and for the way we understand it.

As already discussed, entropy is a concept that is used in a very wide range of applications. The focus of this thesis is the use of entropy in Hydrology. Therefore the first section of the chapter is dedicated to hydrology. The second section presents applications from the natural sciences with a focus on Earth and life sciences. The third section provides one example of the use of entropy to solve an Engineering problem.

The examples of this chapter are not presented in their full detail. The goal is to show how widely entropy is used in hydrology and beyond. Interested readers are encouraged to read the original articles or books where these examples were taken from.

3.1 Entropy and Hydrology

We can subdivide the hydrologic applications of entropy in two categories. In the first category we find applications in the field of hydrometeorology that use entropy within the context of thermodynamics. In the second we find applications in the field of stochastic hydrology that use entropy within the context of maximum entropy statistical inference.

Hydrometeorology

Hydrometeorology is a field of science that integrates elements of hydrology and meteorology to solve problems for which neither hydrology nor meteorology are sufficient (Fry and Showalter, 1945). To a large degree it owes its development to the concept of *Probable Maximum Precipitation* (PMP), which is the alleged upper limit of precipitation of an area. Much of hydrometeorology was developed as scientists tried to find a combination of statistical, hydrologic and meteorological methods to estimate the PMP[17].

[17] I believe that the Probable Maximum Precipitation is a fallacy. When the concept was first developed in the US it was called the *Maximum Possible Precipitation*. For example, according to the American Meteorological Society (1959) the Maximum Possible Precipitation is "*the theoretically greatest depth of precipitation for a given duration that is* physically possible [my emphasis] *over a particular drainage area at a certain time of year*." Soon it was realized that this limit cannot be calculated, only estimated statistically using a very high

In very simplified terms, meteorology is a combination of fluid mechanics with the thermodynamics of the atmosphere.

The thermodynamics of the atmosphere study how the pressure, temperature, energy, etc., of a *packet of air* change as it moves within the atmosphere. For example when a packet of air is forced by the wind to climb a mountain slope, it moves to an area with lower atmospheric pressure, therefore it expands and cools. Therefore it is able to hold less humidity and rain may be caused. This is how the *orographic effect* is caused.

It is through these thermodynamics that entropy comes into the picture within the field of hydrometeorology.

For example *tephigrams* are based on the concept of entropy. They are thermodynamic diagrams of temperature versus entropy with rotated axes. They are usually used to plot air temperature and dew point data from *radiosondes*[18] and to analyze the stability of the atmosphere using the *convective available potential energy* (CAPE) method. Figure 3.1 shows an example of a tephigram with data from a real radiosonde.

return period so the term was renamed. Sometimes this is how science progresses. A new field starts based on an error but along the way correct discoveries are made. The challenge however is to discard the fallacious origins once the field has matured. Hydrometeorology has not yet fully completed this challenge. For more on the critique of the PMP see Koutsoyiannis (2000).

[18] Radiosondes are meteorological probes, usually attached to balloons, that radio transmit their data to a ground station.

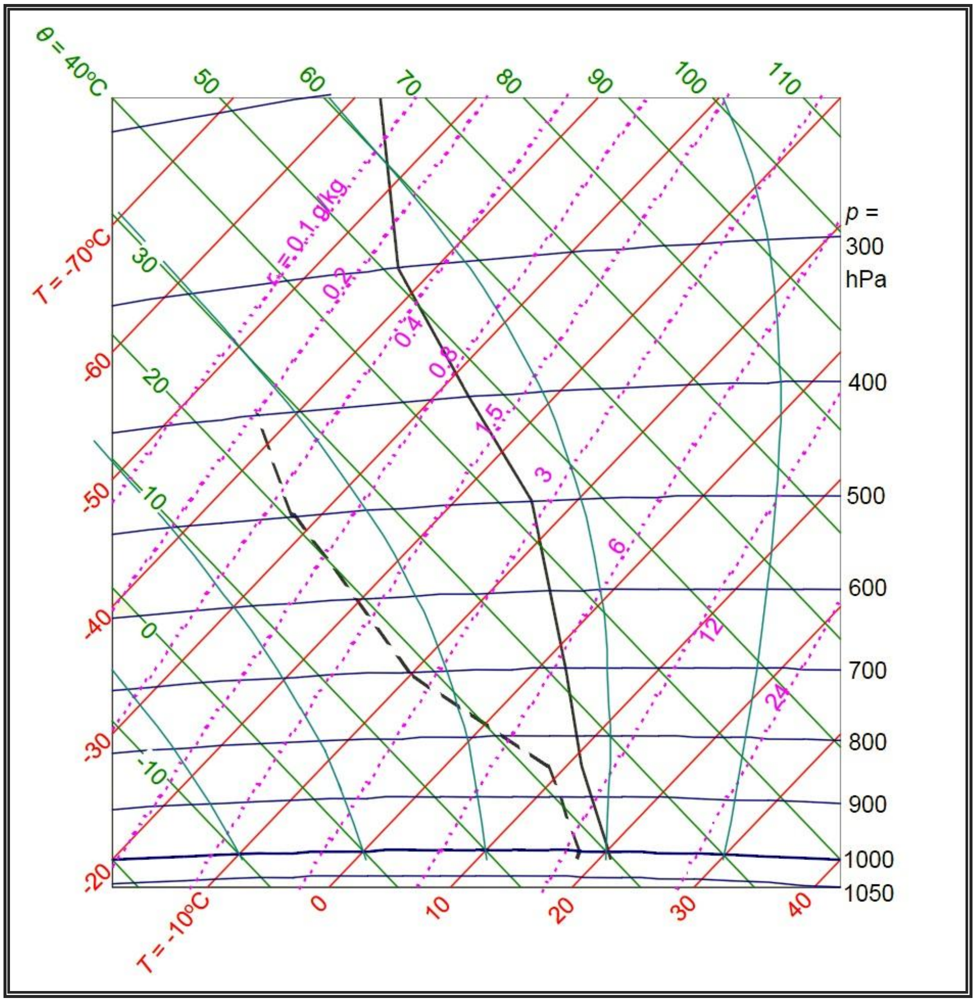


Figure 3.1: Tephigram displaying the temperature (solid black line) and dew point (dashed black line) of a radiosonde conducted by the National Meteorological Agency (EMY) of Greece, in Athens on 5/6/1972 at 00:00 GMT. Source: Koutsoyiannis (2001)

Stochastic Hydrology

Stochastic hydrology is the branch of hydrology that uses the theory of stochastic processes to study hydrologic variables and properties taking into account their probabilistic nature.

Finding the probability distribution function that random hydrologic variables follow is a very important and frequent problem that stochastic hydrology tries to solve. The *Principle of Maximum Entropy* can be applied on this problem.

Claps et al. (1996) studied river geomorphology using entropy. More specifically they compared *river networks* with *mathematical fractals*. Considering the placement of stream segments and junctions as a random process, and using the probability of each segment being placed to a certain position in the network, they define the informational entropy of river networks and of fractals. Then they compare the entropy with other relevant measures, such as the Horton order and the topological diameter. They discovered that these measures are related and concluded that this offers the opportunity to develop new methods to estimate fractal measures of river networks.

Singh (1997, 2000) offers an *extensive review* of how the principle of maximum entropy and other applications of the concept of entropy are used in hydrology by various researchers. He presents applications on the topics of risk and safety from hydrologic system, reliability of water resources systems, hydrologic parameter estimation, water quality, optimization, model selection, hypothesis testing, basin geomorphology etc. He also offers a short but interesting discussion about the significance of the principle of maximum entropy for developing countries, where hydrologic records are limited and therefore it is important to find how to maximize the use of available data.

Kawachi et al. (2001) study the pattern of rainfall in Japan. First they define the *entropy of rainfall*, which is a measure of the spread of rainfall throughout the year. As we have already seen, Shannon's entropy is maximum when events are equiprobable, and zero when one event is certain and the rest are impossible. In analogy, if annual rainfall is equally distributed to all days of a year then entropy is maximum, while if all rainfall of a year falls in one day then entropy is zero. The motivation for the introduction is the fact that "*besides the aggregate rainfall, its temporal apportionment can be a significant aspect of rainfall*" (Kawachi et al., 2001). They used 1107 raingauges to prepare a rainfall *isoentropy map* of Japan. Coupling this map with an isohyetal map they categorized water resources availability using 4 categories that combined the total quantity of rainfall and its annual spread.

Koutsoyiannis (2005a) uses the principle of maximum entropy to show how various *marginal distributions* (such as Gaussian, exponential, Pareto) emerge given various constraints. The example of the Gaussian distribution was already presented in Chapter 2 where the principle of maximum entropy was defined.

Koutsoyiannis (2005b) uses the principle of maximum entropy applied for *joint entropies* to explore the *time dependence* of random variables. He shows that if the dependence is assumed to be dominated by a single time scale, then the Markov process emerges. On the other hand if the dependence is assumed to not be dominated by any time scale, then the Hurst-Kolmogorov process[19] emerges.

Koutsoyiannis (2011b) uses the concept of *entropy production* and the law of maximum entropy production to give a theoretical basis to the use of the Hurst-Kolmogorov process in natural sciences. The entropy production rate of a stochastic process is defined simply as[20] $\Phi'=\mathrm{d}\Phi(X_t)/\mathrm{d}t$. To avoid emergence of infinities he defines *entropy production in logarithmic time*, which is defined as $\varphi'=\mathrm{d}\Phi/\mathrm{d}(\ln t)$. The logarithm is a monotonically increasing function therefore $\Phi(X)\leq\Phi(Y) \Leftrightarrow \varphi(X)\leq\varphi(Y)$, therefore maximum entropy production corresponds to maximum entropy production in logarithmic time. He finds that the Hurst-Kolmogorov process results in extremization of entropy production in asymptotic times, specifically it results in minimum entropy production when time tends to zero and in maximum entropy production when time tends to infinity. These results show that,

> *no other notions (e.g. self-organized criticality, scale invariance, etc.) in addition to entropy extremization are necessary to explain the emergence of the Hurst-Kolmogorov behavior.*

Koutsoyiannis (*Ibid*.) concludes that,

> *extremal entropy production may provide a theoretical background in such stochastic representations* [i.e. the Hurst-Kolmogorov process]*, which otherwise are solely data-driven. A theoretical background in stochastic representations is important also for the reason that* […] *merely statistical arguments do not suffice to verify or falsify the presence of long-term persistence in natural processes.*

From the point of view of hydrologic modeling, the most important conclusion of the last two examples is that when uncertainty is the focus of modeling, for example during the

---

[19] For a definition of the Hurst-Kolmogorov process see Koutsoyiannis (2002), Koutsoyiannis et al. (2011), and Theodoratos (2004)

[20] Koutsoyiannis (2011b) uses the letter Φ to denote entropy.

design of hydropower reservoirs, then the Hurst-Kolmogorov process should be considered, as it is the one that maximizes uncertainty. According to Koutsoyiannis (2011b)

> *The emergence of maximum entropy (i.e., maximum uncertainty) for large time horizons, as demonstrated here, should be considered seriously in planning and design studies, because otherwise the uncertainty would be underestimated and the constructions undersized. The relevance of the last point may be even wider, given the current scientific and public interest on long-term predictions.*

Papalexiou and Koutsoyiannis (2012) use the principle of maximum entropy to search for a theoretically based distribution for daily rainfall. First they explain that despite the availability of long records and significance of daily rainfall, there is still no consensus for the probability distribution of daily rainfall. Then they critique the common method for the selection of distributions, according to which measurements are used to test a number of distributions and select the one with the best fit. As they explain this method is naïve as first, infinite more distributions could theoretically be tested, and second, the final selection is not justified theoretically. Furthermore they explain that using the principle of maximum entropy results in distributions that are uniquely defined by the constraints used. Therefore the selection of constraints becomes the most important part of estimation. They propose three specific constraints. These three constraints yield two distributions that can be used for daily rainfall, the Generalized Gamma (GG, a 3-parameter exponential type) and the Generalized Beta of the 2nd kind (GB2, a 4-parameter power type). They use records from 11 519 rainfall records across the globe and find that the performance of the two distributions is very good. Finally they explain that the great diversity of observed rainfall from very diverse climates can be reproduced by the very flexible GB2 distribution.

3.2 Entropy and Natural Sciences

There are numerous applications of the concept entropy in the natural sciences. In this section I present only a few applications. They are selected, first, to provoke interest in entropy and its implications, and second, to show that the view that entropy is a measure of "disorder" can be misleading.

The concepts of this section are presented in an empirical and qualitative fashion. Furthermore, some analyses and conclusions are based on my understanding of the literature or even on my own intuition, not on scientific proofs. However I feel that they may point towards possible insights regarding the topic at hand which is why I decided to include them in the thesis.

Entropy, Self-Organization and Life

In this section I discuss the application of entropy to the study of self-organized systems, including living organisms and ecosystems. The concept of self-organization has been criticized for lacking an objective definition. For example Koutsoyiannis (2011b) presents as an accomplishment the fact that he manages to explain the emergence of the Hurst-Kolmogorov behavior without using the concept[21] of self-organization. From this it can be inferred that he stands critical towards the use of this concept.

Self-organized systems

A system is *self-organized* when it exhibits some form of global organization that is not imposed externally, but arises out of the system itself. Here non-equilibrium self-organized systems are considered. Many of these systems appear to be violating the second law of thermodynamics if we naïvely view entropy to be related to disorder and if we naïvely forget that the law states that entropy increases in isolated systems as they move towards equilibrium. However non-equilibrium self-organized systems are not isolated. In fact self-organization in such systems can arise only if they are kept out of equilibrium by the flux of energy through the system from some external source.

Example: Benard cells

Benard cells were studied by French physicist *Henri Benard* in 1900[22]. In addition to the cover figure, they can be seen in figure 3.2. They are hexagonal *convection* cells that appear when a thin layer of liquid, for example olive oil, is heated from below, given some favorable conditions. They form as warmer, less dense liquid from the bottom rises, while colder, denser liquid from the top sinks, creating convection currents. Benard observed that sometimes the currents organize themselves in almost hexagonal cells, with liquid rising in the center and sinking around the edges (Ball, 2004).

When the heating is mild there is no convection, heat is transferred to the top slowly by conduction. When the heating is intense, convection currents appear. Currents do not

---

[21] Koutsoyiannis (2011b) uses the word "notion" not the word "concept" which could show that he considers it more as a subjective idea than an objective concept.

[22] Even though they were named after Benard, the convection cells were first observed by German physicist *Georg Hermann Quincke* (Ball, 2004)

always organize themselves in cells. Other patterns can emerge, such as the ones shown in figure 3.3.

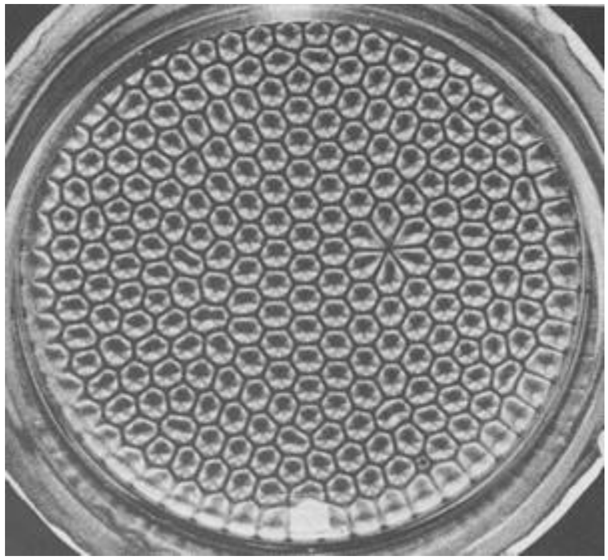


Figure 3.2: An example of Benard cells. Source: Georgia Institute of Technology (www.catea. gatech.edu/grade/mecheng/mod8/mod8.html)

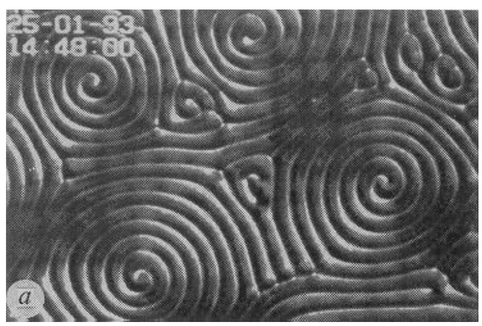


Figure 3.3: Rolling convection currents. Source: Magdeburg University (www.uni-magdeburg.de/ abp/pics/benard-spirale.jpg)

Energy dissipators

Benard cells are an example of *dissipative structures*. According to Ball (2004) dissipative structures are "*organized arrangements in non-equilibrium systems which are dissipating energy and thereby generating entropy*". Dissipative structures are not confined to the dissipation of energy. They are in general dissipating available gradients, such as temperature, pressure, concentration, electric potential, etc. Dissipative structures are omnipresent around us. Vortices and lasers are common examples of such structures. According to Prigogine (Ball, 2004) dissipative structures do not emerge close to equilibrium. But when a system is forced far from equilibrium, i.e. when gradients become steep, systems reach critical points called *bifurcations* which give rise to self-organization.

Example: Turbulence

Turbulence is a flow regime characterized by *stochastic variations* in the velocity and pressure of a flowing fluid. The variations take the form of *vortices*, appearing from very small to large scales. Despite the chaotic nature of the vortices, turbulence is a self-organized phenomenon in a stochastic sense and follows some very simple mathematical laws discovered by Kolmogorov which are described briefly below.

The presence of vortices results in much quicker mixing of properties compared to *laminar flow*, which is the opposite flow regime, where flow is streamlined and lateral mixing occurs only due to molecular diffusion. Figure 3.4 shows these two flow regimes. Flow is laminar when the *Reynolds number* is small (practically, when the flow is slow). When the Reynolds number is high (practically, when the flow is fast) the flow becomes turbulent.

From the point of view of dissipative structures, turbulence can be viewed as an arrangement that accelerates the diffusion of kinetic energy. Kinetic energy can be diffused by viscous forces[23]. But viscous forces become dominant only at small scales. Turbulence establishes a *cascade* of kinetic energy which results to a transfer of energy from large to small scales where it can be diffused to heat. When the flow is slow, in other words when the kinetic energy gradient is mild, laminar flow is sufficient to dissipate energy. But when the flow is fast, in other words when the kinetic energy gradient is steep, the fluid arranges itself in a turbulent flow to increase the dissipation of energy.

[23] Viscous forces of fluids can be viewed as analogous to friction of solids.

Kolmogorov (1941) showed that the smallest scales of turbulence, where the dissipation of kinetic energy to heat becomes dominant, are the same for all flows and depend only on the average rate of dissipation ε and the kinematic viscosity of the fluid ν (Hunt and Vassilicos, 1991). These scales are called *Kolmogorov scales* and follow the equations:

$$\eta = (\nu^3/\varepsilon)^{1/4} \tag{3.1}$$

$$\tau_\eta = (\nu/\varepsilon)^{1/2} \tag{3.2}$$

$$u_\eta = (\nu\varepsilon)^{1/4} \tag{3.3}$$

where η is the smallest length scale, τ_η the smallest time scale, and u_η the smallest velocity scale. He also showed that the *energy spectrum* is a *power law* of the length scale according to the equation:

$$E(k) = C\varepsilon^{2/3}k^{-5/3} \tag{3.4}$$

where E is the kinetic energy, C some constant, and k the length scale (*Ibid.*). The power-law distribution could be related to the fact that kinetic energy is maximized at all scales simultaneously. This maximization leads to faster dissipation. Turbulence is said to be a *statistical fractal* because it follows this power-type scaling law.

The example of turbulence could lead to two important conclusions. First, it could support the idea that self-organization emerges as gradients become steeper. Second, it could show that self-organization can either take a deterministic form, such as hexagonal Benard cells, or a stochastic form, such as chaotic turbulent vortices.

Furthermore it supports the view that the concept of “disorder” is subjective and misleading. Turbulence is not disordered. It may appear so to our human eyes because it is ordered in an uncertain way. But uncertainty does not mean disorder. We may think so because our human minds like certainty and control, and are afraid of uncertainty. But uncertainty makes life beautiful, without it life would be boring. Uncertainty, in the form of turbulence, creates beauty in nature. For example turbulence is what allows water droplets to be suspended in the atmosphere, as the vortices cause stochastic upward movement that lifts the droplets. This allows the formation of clouds. Without turbulence, nature would not be able to give rise to sunsets as the one seen in figure 3.5.

Figure 3.4: Laminar (below) and turbulent (above) flow of cigarette smoke. The turbulent flow leads to significant increase of mixing of smoke with the ambient air. Source: www.cfd-online.com (http://i.minus.com/ie1nQ8.jpg)

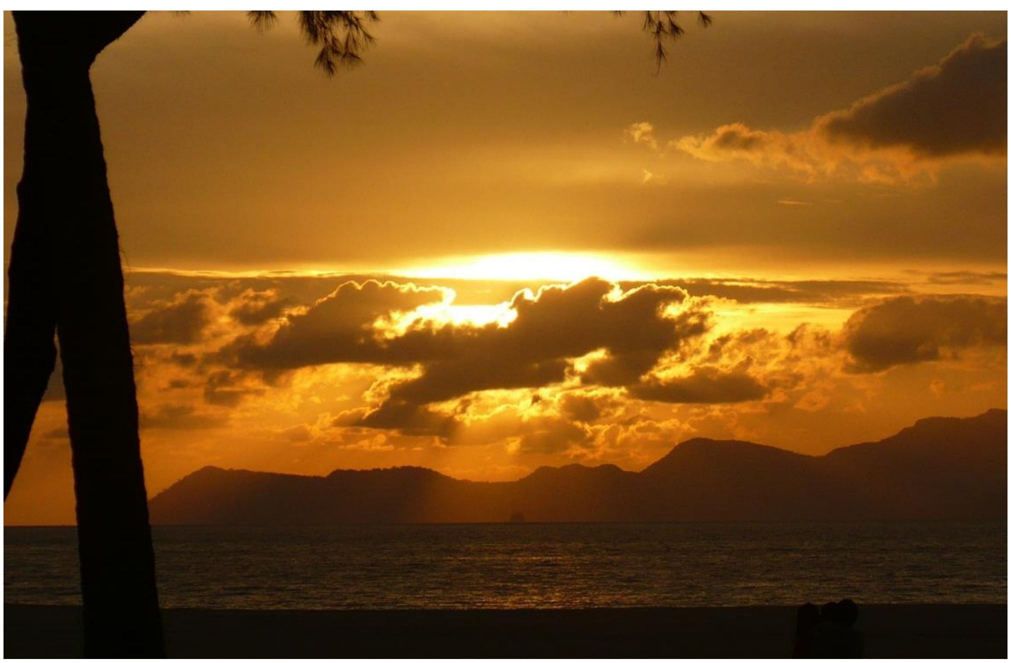


Figure 3.5: Turbulent air flow allows water droplets to be suspended in the atmosphere. Without turbulence clouds would not be able to form. July 2012, Langkawi, Malaysia. Personal photo.

Organization maximizes production

Dissipative structures and self-organization can be explained as a *manifestation* of the *second law of thermodynamics* and the *law of maximum entropy production*. In qualitative terms the explanation can be the following. When a system becomes self-organized it "consumes" entropy, in the sense that its local entropy is reduced. This means that it must produce more entropy than a non-organized system to compensate this local decrease of entropy, if it is to follow the law of maximum entropy production. In other words self-organized systems are more efficient dissipators than non-organized systems. It is important to repeat that self-organization emerges only in non-equilibrium conditions and only in non-isolated systems.

An additional explanation of self-organization is offered by Ozawa et al. (2003). They study turbulent fluid systems in nature such as the atmosphere or the mantle. They show that these systems produce maximum entropy due to thermal and viscous dissipation. They explain that the production of entropy is maximized due to a combination of *positive* and *negative feedback* mechanisms. If the system is initially in a non-organized state, any small random perturbation can initiate a positive feedback loop that leads to the emergence of a self-organized state. On the other hand if the system is in a self-organized state, small perturbations initiate negative feedback loops that lead back to the self-organized state. Thus the self-organized state is a steady state.

The findings of Ozawa et al. (2003) illuminate an additional property of self-organized systems, the property of *homeostasis*. Homeostasis refers to the ability of a system to self-regulate in order to negate changes to its conditions in order to maintain a *steady state*.

Example: Osmotic pressure dissipation through lipid membranes[24]

Lipids are organic molecules which, along with proteins, constitute the membranes of living cells. Lipids have a *hydrophilic* head and a *hydrophobic* tail. In the presence of water, lipids spontaneously self-assemble into *bilayers*, positioning their hydrophobic tails towards the center of the bilayer and the hydrophilic heads facing the aqueous surroundings. This specific arrangement of lipids allows both sides of the membrane to be in contact with water. Furthermore, it makes lipid membranes *semipermeable*. Large hydrophilic molecules cannot be in the interior of the membrane, therefore they cannot travel through the membrane. Water on the other hand can easily travel through membranes due to its

[24] This example is based on personal correspondence with Kamila Oglęcka. At the time of writing the findings presented here were being submitted at a high-impact scientific journal for review.

small molecular size. Therefore at the presence of an osmotic pressure gradient, water can flow through a lipid membrane.

Vesicles are small spheres composed of a lipid bilayer containing aqueous solutions. Compartmentalization of fluids by membranes is a fundamental architecture of living cells, which can be mimicked by artificially creating lipid vesicles that serve as simplified models of cell membranes. Oglęcka is studying the physical properties of synthetic lipid bilayers. She hydrates dried lipid films with sugar-water solutions. This creates vesicles that contain sugar-water solution surrounded by an ambient sugar-water solution of the same concentration. Thus the osmotic gradient across the vesicles' membrane is zero. Dilution of the solution water establishes a concentration gradient across the membranes, creating osmotic pressure. As water flows into the vesicles inflating them, the membrane tension increases until local rupture occurs and sugar solution is released into the ambient bulk. After the release of excess pressure, the membrane rupture heals and a new cycle of inflation-burst starts. The combination of osmosis and release of sugar-water from the vesicles eventually leads to a new equilibrium where the concentration of sugar is approximately equal on both sides of the membrane. This is an example of gradient dissipation phenomena.

Oglęcka has observed the emergence of a self-organized state of vesicle membranes. She uses more than one kind of lipids for the formation of vesicles. Under normal membrane tension lipids of different kinds are mixed. However, under high membrane tension, and given the correct conditions, a *phase separation* in the membrane can take place. Namely different kinds of lipids can become segregated. The separation is visualized by adding a fluorescent dye which prefers to attach itself to one kind of lipids and observing the vesicles using fluorescence microscopy. Images like the one shown in figure 3.6 are observed.

At, or near, equilibrium conditions, that is under the influence of zero, or small, osmotic gradients, lipids of different kinds tend to be fairly uniformly distributed across the membrane. Under the influence of a steeper osmotic gradient, however, a self-organized state emerges. This example follows the same pattern that is being postulated in this section, namely conditions far from equilibrium can lead to self-organization.

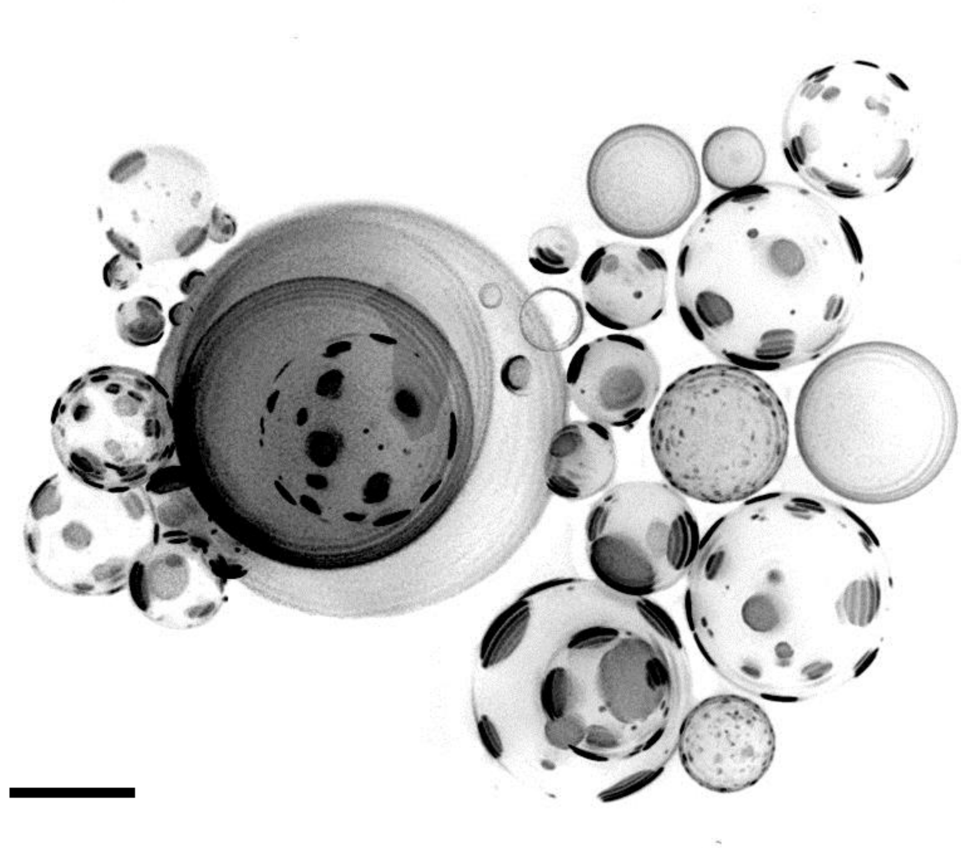


Figure 3.6: Phase separation in lipid membranes of vesicles; an example of self-organization emerging under the influence of an osmotic pressure gradient. Different colors represent different kinds of lipids. The bar in the lower left corner is 10 μm long. Source: Oglęcka (2012).

Living systems

For a long time it was thought that *life* was *defying* the *second law of thermodynamics*. This was due to the naïve view that entropy equals disorder and systems are expected to become disordered as time passes. Living organisms however present remarkable order, while life as a phenomenon leads to ever more order due to *evolution*. This was viewed as a violation of the second law. Therefore biology was seen as a science that cannot be reconciled with physics.

This view is obviously mistaken. Living systems are *not isolated*; they are open, as they exchange both energy and matter with their environment. Organisms stay alive and grow by metabolizing energy and matter that they consume and by depositing high-entropy waste energy and matter to their environment. Furthermore, on a planetary scale, the global ecosystem maintains and evolves itself due to the massive production of entropy by the Sun (Sagan, 2010).

From a different point of view it can be said that life actually increases entropy as it increases uncertainty. For example one can compare the images of Martian landscapes sent to Earth by NASA's Curiosity rover which landed on Mars on August 6, 2012, with images of landscapes from Earth. The uncertainty of how a landscape will look, what forms and shapes or chemical composition it will have is much higher for a landscape on Earth than on barren Mars. More insights along these lines are presented in the next section which deals with entropy and the Earth system.

Living organisms can be viewed as dissipative structures. Their function is to dissipate gradients. For example, trees dissipate the gradient between soil and atmospheric water through the process of transpiration. Or, for example, food is a chemical gradient and it is dissipated through the metabolic processes, by breaking down complex molecules to simpler ones. An organism itself is a gradient, to be dissipated (i.e. eaten) by other organisms.

The *sensors* of organisms can be seen as mechanisms that detect gradients and direct the organism towards the gradient. For example some amoebas can sense changes in the pH of the water and move accordingly. Plants can sense light gradients and can grow towards it to maximize photosynthesis. The eyes of a lion, an example of a much more sophisticated sensor, can detect a zebra, which is a chemical and energy gradient, and direct the lion to try to eat it.

From this perspective evolution can be seen as a process that, amongst other things, optimizes the sensors of organisms so that they can more efficiently detect and dissipate gradients. This perspective can also explain *intentionality*, an additional characteristic of living organisms that seems irreconcilable with physics. Intentionality, in the biologic sense, is the ability of organisms to perform intentional actions, which is in contrast to the unintentional "actions" performed by particles, engines, clouds and rivers.

One of the first to explore life's relation to the concept of entropy was Austrian physicist *Erwin Schroedinger* (2010). The ability of living organisms to maintain themselves, grow and evolve had led many scientists, even in modern times, to claim that there was some unique, meta-physical force in living organisms called *entelechy*. He showed how an organism manages to stay alive only because it is depositing high entropy to its environment.

A very important conclusion is that if dissipative structures emerge due to the law of maximum entropy production and if living organisms are such structures, then *the emergence of life is very probable* when conditions are favorable.

Entropy and the Earth System

The *Earth System* is the integration of the biosphere, atmosphere, hydrosphere and lithosphere in a single system based on the view that processes occurring within and between these subsystems should be studied in a holistic, integrated and interdisciplinary manner, in order to understand and account for their interconnections and feedback mechanisms. The concept of entropy played an important role in the establishment of this point of view.

The concept of the Earth System was developed to a large extent due to the *Gaia hypothesis*. According to this hypothesis the whole biosphere can be seen as a single living being, called *Gaia* (Lovelock, 1979). Each individual organism can be seen as a part of Gaia, having with it the same relation that cells have with an organism. Supporters of the Gaia hypothesis range from those who accept it in a wide sense, as a metaphor that helps in holistic understanding of planetary biogeochemical processes, to those that interpret it strictly and literally. The later have views that reach the limit of science fiction. For example they believe that Gaia can even replicate itself on other planets. Humans can play the role of spores by terraforming and colonizing other planets (Stephen Miller, 1998). Views like this have led to widespread criticism of the hypothesis. However, if used metaphorically, it can lead to a more complete understanding of the Earth System.

The Gaia hypothesis was developed by English chemist *James Lovelock*. During the early 1960's NASA started planning to search whether life existed on Mars. Lovelock believed that first scientists should define what exactly life is, as life on other planets may be very different than life on Earth. He believed that a universal definition should relate life to a local decrease of entropy. For example, Earth's atmosphere contains high concentration of oxygen and methane, which would be impossible under chemical equilibrium conditions. It is made possible by biological processes. Therefore a chemical disequilibrium in a planet's atmosphere, observable remotely by radio telescopes, can be a very strong indication for the presence of life. This idea, along with other observations, led eventually Lovelock to the hypothesis that the biosphere has shaped the atmosphere for its collective benefit, as if the whole biosphere was a single being.

Schwartzman (1999), supporting a metaphorical view of the Gaia hypothesis, showed that the biosphere shapes the lithosphere, in addition to the atmosphere, by the enhancement of *weathering* of silicate minerals by the biosphere. Weathering of silicate minerals is the process of replacement of silicate (SiO_3^{2-}) by carbonate (CO_3^{2-}) in rock minerals according to the reaction

$$CO_2 + CaSiO_3 \rightarrow CaCO_3 + SiO_2$$

Schwartzman uses the concept of entropy and the example of weathering to support the idea that the biosphere as a whole is a self-organizing system, dissipating the energy of the incoming solar radiation.

Kleidon (2009), through a review of the literature, shows that first, the Earth System is a thermodynamic system therefore simulations of global mass and energy transport processes and global biogeochemistry, such as radiative exchange, the carbon and hydrologic cycles, should be based on thermodynamic principles, including the maximum entropy principle. Furthermore he shows that the principle of maximum entropy production proves that the Gaia hypothesis "*may be closer to the truth than what some of its skeptics would expect*". The Gaia hypothesis is interpreted to be explaining that the Earth System has been maintained at a far-from-equilibrium, yet relatively stable, state by and for the benefit of the biosphere.

3.3 Entropy and Engineering

In this section I present an engineering application of the concept of entropy.

Entropy and Measurement Systems

Robert-Nicoud et al. (2005) present an application of the principle of maximum entropy in the problem of measurement system configuration using two examples, one from structural and one from water resources engineering.

A measurement system consists of a number of sensors that measure various properties of a physical system at various locations. Given that the number of sensors is often limited, it is important to optimize the spatial configuration of the sensors so that the data collected can offer a maximum benefit.

Robert-Nicoud et al. (2005) are assuming that the optimal configuration of sensors is the one that allows them to best distinguish between candidate models. They are measuring the separation between models using Shannon's entropy. The entropy is calculated by creating distributions of the values that candidate models predict at the location of each sensor. The locations that lead to higher entropy are the ones that lead to larger separation between models, making it thus easier to select which models are best. They explain that "*the best system is the one in which the total entropy is the maximum. This can be formulated as a discrete optimisation problem.*"

Robert-Nicoud et al. (2005) are offering two examples. The first example is from structural engineering. They present a horizontal timber beam in the laboratory, on which vertical loads are applied. Under the beam are sensors of vertical displacement at various locations. They are searching for the optimal position for a sensor across the beam's length. The second example is from water resources engineering. They study leakage from a water distribution system. Leakage can be detected by sound sensors. They try to identify the optimal installation locations of the sensors.

4. Epilogue

This thesis presented an extended, but far from exhaustive, review of the literature regarding entropy. The main goal was to present the concept of entropy to a hydrological engineering audience and show how it can be applied by hydrologists.

Entropy is one of the most important and at the same most elusive concepts of science. It is usually interpreted as a measure of "disorder". The point of view that I agree with is that this interpretation can be misleading. Therefore a side goal of the thesis was to show that entropy's interpretation as a measure of uncertainty is easier to grasp. Furthermore it is applicable to other areas, such as information theory, where the notion of "disorder" does not seem to make sense.

Entropy was first introduced to give a mathematical form to the second law of thermodynamics but has since found application in a wide range of sciences. The thesis defined entropy first from an information-theory and then from a thermodynamic point of view. Historically however the concept of entropy was developed in the opposite order. Gibbs' definition of thermodynamic entropy had existed for around 50 years when Shannon discovered in 1948 the measure of uncertainty of information theory and realized that it is given by the same equation as Gibbs' entropy. Therefore he called the new measure entropy. More than 60 years later, the debate is still open as to whether this was a wise naming choice and as to whether the two entropies are related or not. The point of view that this thesis defended is that they two entropies are related, as thermodynamic entropy is a special case of Shannon's entropy.

A number of applications of the concept of entropy in hydrology were presented. More specifically the thesis presented applications from hydrometeorology, such as the tephigram. From the area of stochastic hydrology it presented examples of Jaynes' maximum entropy principle, which is a method of statistical inference that can be applied to calculate distributions of hydrological variables. Also, it presented studies of geomorphology and of temporal distribution of rainfall that use the concept of entropy. Finally it presented how entropy can be used to explain the Hurst-Kolmogorov process and the property of persistence, also known as long-term "memory".

Furthermore, special attention was given to applications from other areas of natural sciences to show the breadth and depth of entropy's applicability and provoke the readers' interest. Some of these applications were presented empirically and without rigorous scientific proofs. The first theme of natural sciences that was presented was the possible emergence of self-organized systems due to Ziegler's principle of maximum entropy production rate. Examples of such systems were Benard convective cells, turbulence, osmosis through lipid membranes, and living systems. An important empirical conclusion was that if self-organization is a result of Ziegler's principle and if living systems are an example of self-organized systems, then the emergence of life is very probable when conditions are favorable. The second theme of natural sciences that was presented was the Earth system, which is the integration of the biosphere, atmosphere, hydrosphere and lithosphere in a single system. This view can help promote holistic and interdisciplinary research. It was shown through examples that entropy plays an important role both in the original development of the concept of the Earth system and in current research.

Finally an engineering application of entropy was presented. This application shows how the maximum entropy principle can be used for the design of measurement systems by calculating optimal locations of sensors.

I hope that this document can serve as an easy-to-follow introduction to entropy. However it cannot cover the whole topic. Therefore interested readers are encouraged to study on their own books and publications from the reference list and beyond.

Finally, the concept of entropy can be an interesting and important topic for future hydrologic research. An area of study could be the derivation of distributions of various hydrologic variables using Jaynes' principle of maximum entropy. Another area of study could be turbulence, testing in a rigorous and concrete manner whether turbulence is indeed related to the maximization of entropy production as was speculated in this thesis.

References

Efstratiadis, A., and C. Makropoulos, 2011, *Fundamental concepts of optimization and classical mathematical methods*, in Greek, Notes for the course: Optimization of water resource systems – Hydroinformatics, 19 pages, Department of Water Resources and Environmental Engineering, National Technical University of Athens, Athens

Theodoratos, N., 2004, *Stochastic simulation of two-dimensional random fields with preservation of persistence*, in Greek, Diploma thesis, 69 pages, Department of Water Resources, Hydraulic and Maritime Engineering, National Technical University of Athens, Athens

Koutsoyiannis, D., 1997, *Statistical Hydrology*, 4th edition, in Greek, 312 pages, National Technical University of Athens, Athens

Koutsoyiannis, D., 2000, *Probable Maximum Precipitation*, in Greek, 36 pages, Notes for the course: Hydrometeorology, Department of Water Resources, National Technical University of Athens, Athens

Koutsoyiannis, D., 2001, *Atmosphere Thermodynamics*, in Greek, 51 pages, Notes for the course: Hydrometeorology, Department of Water Resources, National Technical University of Athens, Athens

Koutsoyiannis, D., 2007, *Single variable stationary models*, in Greek, 17 pages, Notes for the course: Stochastic Methods in Water Resources, Department of Water Resources, National Technical University of Athens, Athens

American Meteorological Society, 1959, *Glossary of Meteorology*, 638 pages, Boston

Atkins, P., 2010, *The laws of thermodynamics: A very short introduction*, 103 pages, Oxford University Press, New York

Ball, P., 2004, *Critical mass. How one thing leads to another*, 644 pages, Arrow Books, London

Ben-Naim, A., 2008, *Statistical Thermodynamics Based on Information. A Farewell to Entropy*, 384 pages, World Scientific Publishing, Singapore

Borel, L., and D. Favrat, 2010, *Thermodynamics and energy systems analysis. From energy to exergy*, 795 pages, EPFL Press, Italy

Brown, H. R., W. Myrvold, and J. Uffink, 2009, Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics, *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 40(2), 174-191

Callender, C., 1999, Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy, *The Journal of Philosophy*, 96(7), 348-373

Cengel, Y. A., and M. A. Boles, 2010, *Thermodynamics An Engineering Approach*, 5th edition, 881 pages, McGraw-Hill

Claps, P., M. Fiorentino, and G. Oliveto, 1996, Informational entropy of fractal river networks, *Journal of Hydrology*, 187, 145-156

Dugdale, J. S., 1996, *Entropy and its Physical Meaning*, 198 pages, Taylor and Francis, London

Evans, D. J., and D. J. Searles, 2002, The Fluctuation Theorem, *Advances in Physics*, 51(7), 1529-1585

Fry, A. S., and A. K. Showalter, 1945, Hydrometeorology, in Berry, F. A., E. Bollay, and N. R. Beers, (ed.), *Handbook of Meteorology*, 1068 pages, McGraw-Hill, New York

Hunt, J. C. R., and J. C. Vassilicos, 1991, Kolmogorov's contributions to the physical and geometrical understanding of small-scale turbulence and recent developments, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 434, 183-210

Jaynes, E. T., 1957, Information Theory and Statistical Mechanics, *The Physical Review*, 106(4), 620-630

Kawachi, T., T. Maruyama, and V. P. Singh, 2001, Rainfall entropy for delineation of water resources zones in Japan, *Journal of Hydrology*, 246, 36-44

Kleidon, A., 2009, Nonequilibrium thermodynamics and maximum entropy production in the Earth system, *Naturwissenschaften*, 96, 653-677

Kleidon, A., Y. Malhi, and P. M. Cox, 2010, Maximum entropy production in environmental and ecological systems, *Philosophical Transactions of the Royal Society Biological Sciences*, 365, 1297-1302

Kolmogorov, A. N., 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, translation, 1950, *Foundations of the Theory of Probability*, 2nd edition, 84 pages, Chelsea Publishing Company, New York

Kolmogorov, A. N., 1941, *Dokl. Akad. Nauk SSSR*, translation, 1991, Dissipation of Energy in the Locally Isotropic Turbulence, *Proceedings of the Royal Society: Mathematical and Physical Sciences*, 434, 15-17

Koutsoyiannis, D., 2002, The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47(4), 573-595

Koutsoyiannis, D., 2005, Uncertainty, entropy, scaling and hydrological stochastics. 1. Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50(3), 381-404

Koutsoyiannis, D., 2005, Uncertainty, entropy, scaling and hydrological stochastics. 2. Time dependence of hydrological processes and time scaling, *Hydrological Sciences Journal*, 50(3), 405-426

Koutsoyiannis, D., 2011, *A probability-based introduction to atmospheric thermodynamics – determinism free – self-contained*, Lecture notes on Hydrometeorology, Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Athens

Koutsoyiannis, D., 2011, Hurst-Kolmogorov dynamics as a result of extremal entropy production, *Physica A: Statistical Mechanics and its Applications*, 390 (8), 1424-1432

Koutsoyiannis, D., A. Paschalis, and N. Theodoratos, 2011, Two-dimensional Hurst-Kolmogorov process and its application to rainfall fields, *Journal of Hydrology*, 398 (1-2), 91-100

Lovelock, 1979, *Gaia A new look at life on Earth*, Oxford University Press

Martyushev, L. M., 2010 The maximum entropy production principle: two basic questions, *Philosophical Transactions of the Royal Society Biological Sciences*, 365, 1333-1334

Martyushev, L. M., and V. D. Seleznev, 2006, Maximum entropy production principle in physics, chemistry and biology, *Physics Reports*, 426, 1-45

Miller, S., 1998, Is the Human Species the Organ of Gaian Reproduction?, *Gaianation.net*, (http://www.gaianation.net/org/gaiasporing.html)

Moran, M. N., and H. N. Shapiro, 2000, *Fundamentals of engineering thermodynamics*, 4th edition, 918 pages, Wiley, New York

Oglecka, K., 2012, Personal correspondence

Ozawa, H., A. Ohmura, R. D. Lorenz, and T. Pujol, 2003, The Second Law Of Thermodynamics And The Global Climate System: A Review Of The Maximum Entropy Production Principle, *Reviews of Geophysics*, 41(4), 1018-2003

Papalexiou, S.M., and D. Koutsoyiannis, 2012, Entropy based derivation of probability distributions: A case study to daily rainfall, *Advances in Water Resources*, 45, 51–57

Poe, E. A., 1843, e-book in http://www.eapoe.org/works/tales/goldbga2.htm

Price, R., 1982, Interview with Claude Shannon, electronic version available from IEEE, in http://www.ieeeghn.org/wiki/index.php/Oral-History:Claude_E._Shannon

Robert-Nicoud, Y., B. Raphael, and I. F. C. Smith, 2005, Configuration of measurement systems using Shannon's entropy function, *Computers and Structures*, 83, 599-612

Sagan, C., 2010, Definitions of life, pages 303-306 in Bedau, M. A., and C. E. Cleland, (ed), *The Nature of Life Classical and Contemporary Perspectives from Philosophy and Science*, 418 pages, Cambridge University Press, electronic version available in http://ebooks.cambridge.org/ebook.jsf?bid=CBO9780511730191

Schroedinger, E., 2010, What is Life?, pages 50-69 in Bedau, M. A., and C. E. Cleland, (ed), *The Nature of Life Classical and Contemporary Perspectives from Philosophy and Science*, 418 pages, Cambridge University Press, electronic version available in http://ebooks.cambridge.org/ebook.jsf?bid=CBO9780511730191

Schwartzman, D., 1999, *Life, temperature, and the earth: the self-organizing biosphere*, 241 pages, Columbia University Press, New York, USA

Shannon, C. E., 1948, A Mathematical Theory of Communication, *Mobile Computing and Communications Review*, 5(1), 3-55

Singh, V. P., 1997, The Use Of Entropy In Hydrology And Water Resources, *Hydrological Processes*, 11, 587-626

Singh, V. P., 2000, The entropy theory as a tool for modelling and decisionmaking in environmental and water resources, *Water SA*, 26(1), 1-11

Tribus, M., and E. C. McIrvine, 1971, Energy and Information, *Scientific American*, 225, 179-188

Wicken, J. S., 1987, Entropy and Information: Suggestions for Common Language, *Philosophy of Science*, 54(2), 176-193

Websites:

www.cfd-online.com, http://i.minus.com/ie1nQ8.jpg

Georgia Tech, http://www.catea.gatech.edu/grade/mecheng/mod8/mod8.html

liamscheff.com, http://liamscheff.com/wp-content/uploads/2010/07/convection-cells- closeup.jpg

Appendix A

This Appendix presents additional definitions and properties from probability theory[25].

Joint, marginal, and conditional distributions

For two random variables, X and Y, we can define the *joint cumulative distribution function*

$$F_{XY}(x\,,y) = \mathrm{P}(X \leq x\,,\, Y \leq y) \tag{A.1}$$

If F_{XY} is doubly derivable, we can define the *joint probability density function*

$$f_{XY}(x\,,y) = \frac{\partial^2 F_{XY}(x\,,y)}{\partial x \partial y} \tag{A.2}$$

By inversion we get

$$F_{XY}(x\,,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(\xi\,,\psi)d\psi d\xi \tag{A.3}$$

The *marginal cumulative distribution functions* of X and Y are

$$F_X(x) = P(X \leq x) = \lim_{y\to\infty} F_{XY}(x\,,y) \tag{A.4}$$

$$F_Y(y) = P(Y \leq y) = \lim_{x\to\infty} F_{XY}(x\,,y) \tag{A.5}$$

and the *marginal probability density functions* of X and Y are

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x\,,y)dy \tag{A.6}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x\,,y)dx \tag{A.7}$$

The *conditional cumulative distribution function* of X given the value of Y is

$$F_X(x \mid Y = y) = \frac{\int_{-\infty}^{x} f_{XY}(\xi\,,y)d\xi}{f_Y(y)} \tag{A.8}$$

[25] Equations and definitions of this subsection are according to Koutsoyiannis (1997) and Theodoratos (2004), unless otherwise specified

and the *conditional probability density function* of X given the value of Y is

$$f_X(x \mid Y = y) = \frac{f_{XY}(x\,,y)}{f_Y(y)} \qquad \text{(A.9)}$$

For *independent* random variables we have

$$F_{XY}(x\,,y) = F_X(x)F_Y(y) \qquad \text{(A.10)}$$

and

$$f_{XY}(x\,,y) = f_X(x)f_Y(y) \qquad \text{(A.11)}$$

Expected values, moments

Expected values and moments are values that give important information about the average magnitude of a random variable and the shape of its distribution. The *expected value* of a function is simply a probability-weighted average of this function. The *moments* are expected values of a certain class of functions. The most well-known moments are the mean (often refered to as average) and the variance.

If X is a random variable and $g(X)$ a function of the random variable then we define the *expected value* E[$g(X)$], for continuous random variables by the equation

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \qquad \text{(A.12)}$$

and for discrete random variables by the equation

$$\mathrm{E}[g(X)] = \sum_{i=1}^{\infty} g(x)P_X(x_i) \qquad \text{(A.13)}$$

There is a number of special functions $g(X)$ whose expected functions are frequently used. These functions and the respective expected values are presented below.

1. For $g(X) = X^r$ with r = 1, 2, …, we define the moment of order r about zero of X or raw moment of order r of X:

$$m_X^{(r)} = \mathrm{E}[X^r] \qquad \text{(A.14)}$$

2. For $g(X) = X$ we define the *expected value* of X or *mean* of X or raw moment of order 1:

$$m_X = \mathrm{E}[X] \qquad \text{(A.15)}$$

3. For $g(X) = (X - m_X)^r$ with r = 1, 2, …, we define the *central moment of order r* of X:

$$\mu_X^{(r)} = \mathrm{E}[(X - m_X)^r] \qquad \text{(A.16)}$$

4. For $g(X) = (X - m_X)$, the expected value of $g(X)$ – i.e. the central moment of order 1 - is always equal to zero:

$$E[(X - m_X)] = \mu_X^{(1)} = 0 \tag{A.17}$$

5. For $g(X) = (X - m_X)^2$, we define the *variance* of X or central moment of order 2:

$$\sigma_X^2 = \mathrm{Var}[X] = \mu_X^{(2)} = E[g(X)] = E[(X - m_X)^2] \tag{A.18}$$

Based on the variance of X we define two coefficients that are used often. These coefficients are the *standard deviation*:

$$\sigma_X = \sqrt{\mathrm{Var}[X]} \tag{A.19}$$

and the coefficient of variation:

$$C_{Vx} = \frac{\sigma_X}{m_X} \tag{A.20}$$

The definition of expected values can be extended to joint and conditional distributions of two random variables. Thus for a function $g(X, Y)$ of two continuous random variables X and Y we have

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\,dy\,dx \tag{A.21}$$

and for a function of two discrete random variables X and Y we have

$$E[g(X, Y)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x, y) P_{XY}(x_i, y_j) \tag{A.22}$$

Some special expected values of X and Y are:

6. $E[X^pY^q]$ called joint raw moment of order p + q of X and Y.

7. $E[(X - m_X)^p(Y - m_Y)^q]$, called *joint central moment of order $p + q$* of X and Y (m_X and m_Y are the marginal expected values of X and Y).

8. The most frequently used joint central moment is of order 1+1, is called *covariance* of X and Y, and is given by the equation

$$\sigma_{XY} = \mathrm{Cov}[X, Y] = E[(X - m_X)(Y - m_Y)] = E[XY] - m_X m_Y \tag{A.23}$$

Based on the covariance we define the *coefficient of correlation* of X and Y, according to the equation:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \tag{A.24}$$

The coefficient of correlation is dimensionless and takes values $-1 \leq \rho_{XY} \leq 1$. It is a very useful parameter when we study the correlation of two random variables.

It is easily proven that if two random variables are independent then they are uncorrelated, i.e. their covariance is equal to zero. The opposite is not always true, i.e. two random variables may be uncorrelated but dependent.

For the proof we start from the definition of covariance (equation A.23)

$$\sigma_{XY} = \mathrm{E}[(X - m_X)(Y - m_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)(y - m_Y)f_{XY}(x\,,y)dydx$$

According to the property of equation A.11 for independent variables this becomes

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)(y - m_Y)f_X(x)\,f_Y(y)dydx \;\; = \int_{-\infty}^{\infty}(x - m_X)f_X(x)dx\int_{-\infty}^{\infty}(y - m_Y)f_Y(y)dy \;\; = 0 \qquad \text{(A.25)}$$

since these are the central moments of order 1 which according to equation A.17 are zero. For independent random variables we get from equation A.23

$$0 = \sigma_{XY} = \mathrm{E}[XY] - m_X m_Y$$

i.e. for independent variables

$$\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y] \qquad \text{(A.26)}$$

Appendix B

We are trying to prove that for a given distribution p_i, i = 1, 2, ..., n, uncertainty H is maximum when all the events i are equiprobable, and have probabilities

$$p_i = \frac{1}{n} \tag{B.1}$$

using the method of *Lagrange multipliers* (Efstratiadis and Makropoulos, 2011).

We are searching for the maximum of the function

$$H(p_1, p_2, ..., p_n) = - \sum_{i=1}^{n} p_i \ln p_i \tag{B.2}$$

Subject to the constraint

$$\sum_{i=1}^{n} p_i = 1 \tag{B.3}$$

We define the auxiliary function

$$\varphi(p_i, \lambda) = H(p_1, p_2, ..., p_n) - \lambda \sum_{i=1}^{n} p_i \tag{B.4}$$

We calculate the partial derivatives of φ with respect to each p_i,

$$\left(\frac{\partial \varphi}{\partial p_i}\right) = - \ln p_i - 1 + \lambda \tag{B.5}$$

and demand that they are equal to zero

$$- \ln p_i - 1 + \lambda = 0 \Leftrightarrow$$

$$p_i = e^{\lambda - 1} \tag{B.6}$$

Substituting this in constraint equation (B.3) yields

$$1 = \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} e^{\lambda-1} = e^{\lambda-1} \sum_{i=1}^{n} 1 = n e^{\lambda-1} = n p_i \tag{B.7}$$

Therefore

$$p_i = \frac{1}{n} \tag{B.8}$$

Appendix C

This appendix presents extensive quotes from Jaynes (1957), where he introduced the principle of maximum entropy. These quotes offer a number of observations and ideas that can lead to very interesting philosophical implications about maximum entropy inference; the laws of nature and our knowledge of them; and the meaning of probabilities. The quotes are presented without comments. Some of them are not referring directly to entropy, but they show how entropy is related to various other concepts and themes of physics and philosophy.

From the abstract, page 620, regarding experimental verification of results of the principle of maximum entropy:

> *In the resulting "subjective statistical mechanics" the usual rules are thus justified independently of any physical argument, and in particular independently of experimental verification; whether or not the results agree with experiment, they still represent the best estimates that could have been made on the basis of the information available.*

From page 622, regarding the meaning of probabilities:

> *Probability theory has developed in two very different directions as regards fundamental notions. The "objective" school of thought regards the probability of an event as an objective property of that event, always capable in principle of empirical measurement by observation of frequency ratios in a random experiment. In calculating a probability distribution the objectivist believes that he is making predictions which are in principle verifiable in every detail, just as are those of classical mechanics.* […]
>
> *On the other hand, the "subjective" school of thought regards probabilities as expressions of human ignorance; the probability of an event is merely a formal expression of our expectation that the event will or did occur, based on whatever information is available. To the subjectivist, the purpose of probability theory is to help us in forming plausible conclusions in cases where there is not enough information available to lead to certain conclusions; thus detailed verification is not expected.* […]
>
> *Although the theories of subjective and objective probability are mathematically identical, the concepts themselves refuse to be united. In the various statistical problems presented to us by physics, both viewpoints are required. Needless controversy has resulted from attempts to uphold one or the other in all cases. The*

> *subjective view is evidently the broader one, since it is always possible to interpret frequency ratios in this way; furthermore, the subjectivist will admit as legitimate objects of inquiry many questions which the objectivist considers meaningless. The problem posed at the beginning of this section* [finding the distribution of pi of equation 2.74] *is of this type, and therefore in considering it we are necessarily adopting the subjective point of view.*

From page 624, regarding the laws of physics and equilibrium states:

> *There is nothing in the general laws of motion that can provide us with any additional information about the state of a system beyond what we have obtained from measurement.*

The whole paragraph containing this quote is the following:

> *It is interesting to note the ease with which these rules of calculation* [of thermodynamic quantities derived using the principle of maximum entropy] *are set up when we make entropy the primitive concept. Conventional arguments, which exploit all that is known about the laws of physics, in particular the constants of the motion, lead to exactly the same predictions that one obtains directly from maximizing the entropy. In the light of information theory, this can be recognized as telling us a simple but important fact:* there is nothing in the general laws of motion that can provide us with any additional information about the state of a system beyond what we have obtained from measurement [Jaynes' emphasis]*. This refers to interpretation of the state of a system at time t on the basis of measurements carried out at time t. For predicting the course of time-dependent phenomena, knowledge of the equations of motion is of course needed. By restricting our attention to the prediction of equilibrium properties as in the present paper, we are in effect deciding at the outset that the only type of initial information allowed will be values of quantities which are observed to be constant in time. Any prior knowledge that these quantities would be constant (within macroscopic experimental error) in consequence of the laws of physics, is then redundant and cannot help us in assigning probabilities.*

From page 626, regarding the meaning of entropy:

> *Entropy as a concept may be regarded as a measure of our degree of ignorance as to the state of a system; on the other hand, for equilibrium conditions it is an experimentally measurable quantity, whose most important properties were first found empirically. It is this last circumstance that is most often advanced as an argument against the subjective interpretation of entropy.*

From page 626, regarding the justification of maximum entropy inference:

> *One might then ask how such probabilities could be in any way relevant to the behavior of actual physical systems. A good answer to this is Laplace's famous remark that probability theory is nothing but "common sense reduced to calculation." If we have little or no information relevant to a certain question, common sense tells us that no strong conclusions either way are justified. The same thing must happen in statistical inference, the appearance of a broad probability distribution signifying the verdict, "no definite conclusion." On the other hand, whenever the available information is sufficient to justify fairly strong opinions, maximum- entropy inference gives sharp probability distributions indicating the favored alternative. Thus, the theory makes definite predictions as to experimental behavior only when, and to the extent that, it leads to sharp distributions.*
>
> *When our distributions broaden, the predictions become indefinite and it becomes less and less meaningful to speak of experimental verification. As the available information decreases to zero, maximum-entropy inference (as well as common sense) shades continuously into nonsense and eventually becomes useless. Nevertheless, at each stage it still represents the best that could have been done with the given information.*

From page 627, regarding the existence of macroscopic properties of matter and the existence of physics that describe them:

> *Evidently, such sharp distributions for macroscopic quantities can emerge only if it is true that for each of the overwhelming majority of those states to which appreciable weight is assigned, we would have the same macroscopic behavior. We regard this, not merely as an interesting side remark, but as the essential fact without which statistical mechanics could have no experimental validity, and indeed without which matter would have no definite macroscopic properties, and experimental physics would be impossible. It is this principle of "macroscopic uniformity" which provides the objective content of the calculations, not the probabilities per se.*

From page 627, regarding the discovery of new laws of physics:

> *Consider now the case where the theory makes definite predictions and they are not borne out by experiment.* [...] *The most reasonable conclusion in this case is*

> *that the enumeration of the different possible states (i.e., the part of the theory which involves our knowledge of the laws of physics) was not correctly given. Thus, experimental proof that a definite prediction is incorrect gives evidence of the existence of new laws of physics. The failures of classical statistical mechanics, and their resolution by quantum theory, provide several examples of this phenomenon.*

From page 629, regarding the meaning of entropy:

> *We accept the von Neumann-Shannon expression for entropy, very literally, as a measure of the amount of uncertainty represented by a probability distribution; thus entropy becomes the primitive concept with which we work, more fundamental even than energy.*

From page 630, regarding the avoidance of bias by the maximum entropy principle:

> *In the problem of prediction, the maximization of entropy is not an application of a law of physics, but merely a method of reasoning which ensures that no unconscious arbitrary assumptions have been introduced.*