

Was ist Bewusstsein: Integrated Information Theory



Torben Schüngel
21.01.2021

Inhaltsverzeichnis

- Leitfragen
- Existenz von Bewusstsein
- Erfahrung
- Integrated Information Theory (IIT)
- Zusammenfassung und Diskussionsfragen

Leitfragen

Fragen, deren Antworten Struktur vorgeben

Zugrundeliegende Frage: Was ist Bewusstsein?

Dazu mehrere mögliche Unterfragen:

- Wo?
- Wie?
- Warum?

Existenz von Bewusstsein

Konzept: Hier

- “Cogito ergo sum” - Ich denke, also bin ich, René Descartes
- Grundsatz, der für einen persönlich genügt um Bewusstsein in sich selbst zu beweisen
- Jedoch, Basis sehr limitiert. Selbst eigener Körper wäre abgeleitet
- Unsicherheit, ob unser Bewusstsein nicht woanders ist als unsere Wahrnehmung
- Warum nur bestimmte Teile des Gehirns?
 - Die Großhirnrinde ist der Bereich für Bewusstsein obwohl das Kleinhirn mehr Neuronen hat
- Können Computer Bewusstsein haben, oder können sie sich nur verhalten wie wir, allerdings ohne Bewusstsein?

Existenz von Bewusstsein

Konzept: Dort

- Normalerweise, Bereitschaft anderen Bewusstsein zuzugestehen, wenn sie von Gefühlen sprechen
- Beliebte Tests: Erzwungene Binäre Entscheidungen, zum Beispiel “Verärgertes Gesicht vs. Freundliches Gesicht”
 - Unter anderem gesunde Erwachsene aber auch neuronale Netze in der Lage Unterscheidung durchzuführen
 - Sind neuronale Netze daher bewusst?
- Aber mit Menschen mit Gehirnschäden? Oder Säuglinge?
 - Sollte Kommunikationsfähigkeit das einzige Kriterium sein?
- Auf der anderen Seite, Schlafwandeln
 - Sprechen und gehen möglich, aber werden Erfahrungen gemacht?
- Schlüsselkonzept: Erfahrung

Existenz von Bewusstsein

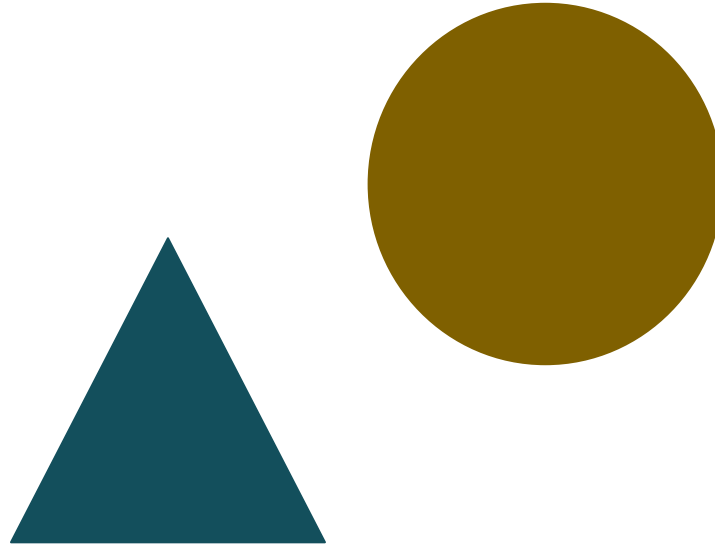
Konzept: Neurale Korrelationen von Bewusstsein (NKB)

- Definition: NKB sind minimale neurale Mechanismen, die zusammen ausreichen für irgendeine/n bewusste Wahrnehmung, Gedanken oder Erinnerung unter konstanten Hintergrundbedingungen
- Hintergrundbedingungen: ermöglichende Faktoren, unter Anderem Sauerstoffversorgung des Gehirns und viele weitere
- Jede Erfahrung hat zugehörige NKB
- Was ist eine Erfahrung?

Beispiel einer Erfahrung

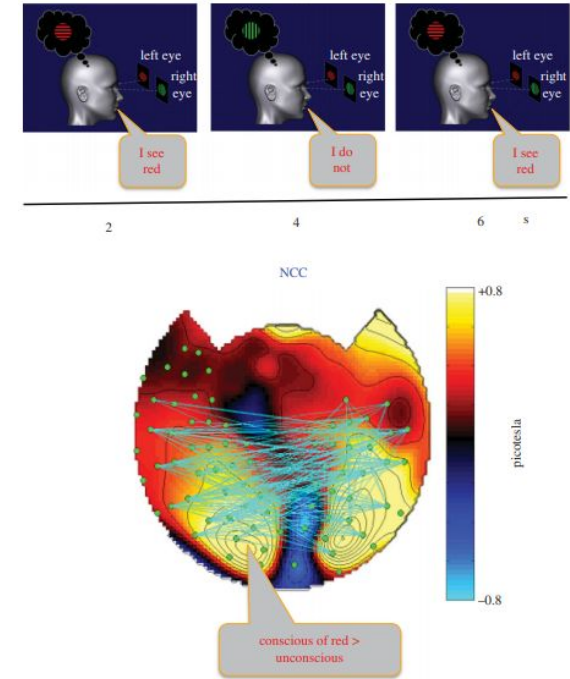


Weiteres Beispiel einer Erfahrung



Erfahrung

- Oben, Verhaltens Korrelationen von Bewusstsein (VKB, engl. BCC für Behavioural Correlates of Consciousness)
- Linkes Auge sieht rot, rechtes grün. Wechsel des Auges jede Sekunde
- Erfahrung wird über Kommunikation erfragt (Sprache, Knöpfe, etc.)
- Unten NKB (engl. NCC für Neural correlates of consciousness), Messung mithilfe von Magnetoenzephalographie, unterschiedliche Frequenzen für rot und grün



Abbildungen: Giulio Tononi und Cristof Koch
Consciousness: here, there and everywhere?, Figure 1

Erfahrung

- Jede Erfahrung hat zugehörige NKB
- Hirnaktivität kann gemessen werden
 - z.B. mit Magnetresonanztomographie (MRT) oder Elektroenzephalografie (EEG)
- Unterschiedliche Hirnresponsen können indikativ für Bewusstsein sein, zum Beispiel:
 - Starke Aktivierung von Frontal- und Parietallappen, Hochfrequente, elektrische Aktivität in Gamma Bereich oder EEG Event P300 Welle
 - Jedoch, Bewusstsein auch ohne Frontallappenaktivität, Gamma Aktivität auch ohne Bewusstsein oder Bewusstsein auch ohne P300, zum Beispiel beim Träumen
- Außerdem, Gehirnaktivität nötig um Fähigkeit zur bewussten Wahrnehmung zu haben
 - Das wäre keine Erfahrung, sondern eine Hintergrundbedingung
- Möglicherweise, andere NKB in Patienten mit Hirnschäden, Säuglingen oder Tieren
- Maschinen wiederum sehr anders

Integrated Information Theory

- Für formale Herangehensweise, Notwendigkeit des LöSENS von VKB und NKB
- Frage nach “Wo ist Bewusstsein?” zweitrangig
- “Wie” und “Warum” jetzt im Vordergrund
- Ansatz: Start bei Bewusstsein, Identifizierung der Eigenschaften und dann Frage nach Art der physikalischen Mechanismen
- Dafür, zunächst Identifizierung von Axiomen der Erfahrung, danach Formulierung von zugehörigen Postulaten

Integrated Information Theory (IIT)

Axiome

- Intrinsische Existenz, Bewusstsein *existiert*, meine Erfahrung *ist*
- Komposition, jede Erfahrung besteht aus *phänomenologischen Unterscheidungen*
 - Beispiel: Innerhalb der selben Erfahrung: Mögliche Unterscheidung eines Buches, der Farbe blau, eines blauen Buches und so weiter
- Information, Bewusstsein ist *spezifisch*, eine Erfahrung unterscheidet sich von jeder anderen möglichen Erfahrung
 - Beispiel: Absolute Dunkelheit unterscheidet sich von anderen möglichen Erfahrungen durch nichtvorhandensein von Formen, Geräuschen, Farben, etc.

Integrated Information Theory (IIT)

Axiome

- Integration, Bewusstsein ist geeint, eine Erfahrung kann nicht in zu zueinander unabhängigen Subsets reduziert werden
 - Gesamte Szene wird wahrgenommen, nicht nur die linke oder rechte Hälfte (Split-Brain Patienten: zwei Instanzen von Bewusstsein, daher Erfahrungen müssen nicht übereinstimmen)
 - Beispiel: Wort Blauwal, blau existiert hier nicht alleine. Beide zusammen ergeben Erfahrung
- Exklusion, Bewusstsein ist definit, jede Erfahrung hat phänomenale Unterscheidung die sie hat, nicht mehr, nicht weniger
 - Beispiel: Wahrnehmung von Bücherregal mit Büchern, darunter ein blaues Buch. Wahrnehmung kann nicht ohne Unterscheidung “blau/nicht-blau” stattfinden.
 - Ebenso nehme ich nicht mehr wahr, zum Beispiel meinen Blutdruck

Integrated Information Theory (IIT)

Postulate

- Intrinsische Existenz, System von Mechanismen muss Ursache-Wirkung Vermögen besitzen, Annahme, dass etwas existiert, wenn nichts dieses beeinflussen kann oder nichts durch dieses beeinflusst wird zwecklos
 - Außerdem, etwas, das in eigener, intrinsischer Perspektive existiert, muss Ursache-Wirkung Vermögen auf selbst besitzen
- Komposition, System muss strukturiert sein
 - Beispiel: Ein System ABC besteht aus Elementen A, B und C. Jedes Subset von Elementen, also A, B, C, AB, AC, BC und ABC kann Mechanismus bilden, der Ursache-Wirkung Vermögen besitzt

Integrated Information Theory (IIT)

Postulate

- Information, System muss Ursache-Wirkung Struktur spezifizieren, die sich von anderen unterscheidet. Das Ursache-Wirkung Repertoire charakterisiert Ursache-Wirkung Vermögen des Mechanismus
- Integration, Ursache-Wirkung Struktur muss einheitlich sein, also intrinsisch Unreduzierbar zu nicht-interabhängigen Sub-Sytemen
- Exklusion, Ursache-Wirkung Struktur muss definit sein, spezifiziert über ein einzelnes Set von Elementen, weder weniger noch mehr

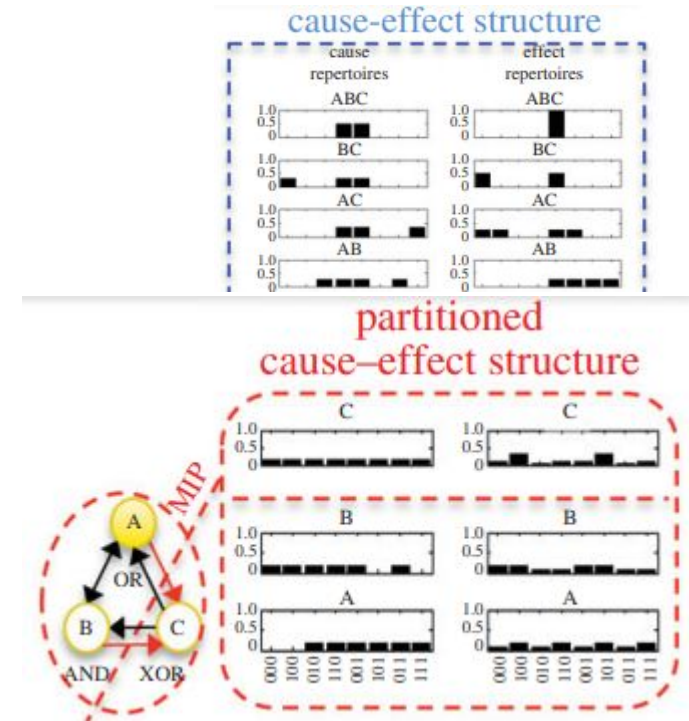
Integrated Information Theory (IIT)

Einige Größen

- Vier Größen werden zur Quantifizierung verwendet:
 - ϕ Integrierte Information, Angabe der Information, die Information der kleinsten Elemente übersteigt
 - ϕ^{\max} integrierte Information eines Konzepts (Mechanismus und zugehörige maximal-unreduzierbares Ursache-Wirkung Repertoire)
 - Φ Integrierte, konzeptionelle Information, beschreibt konzeptionelle Information, die die von Elementen übersteigt
 - Φ^{\max} lokales Maximum der integrierten, konzeptionellen Information. Beschreibt Komplex

Integrated Information Theory

- Gezeigtes System hat Ursache-Wirkung Repertoire
- Sub-systeme (erreicht durch Trennung einzelner Verbindungen) haben anderes Ursache-Wirkung Repertoire
- Prinzip: Minimale Informations Partition (MIP), schwächste Verbindung wird getrennt
- Je “Unreduzierbarer”, desto höher Integrierte Information Φ
- Φ größer, je mehr Veränderung bei Partition



Abbildungen: Giulio Tononi und Cristof Koch
Consciousness: here, there and everywhere?, Figure 3

Integrated Information Theory

- Ursache-Wirkung Struktur in Komplex
- Drei Logikgatter, in Abbildung in Status ABC=100
- Je nach Ursache, unterschiedliche Wirkung

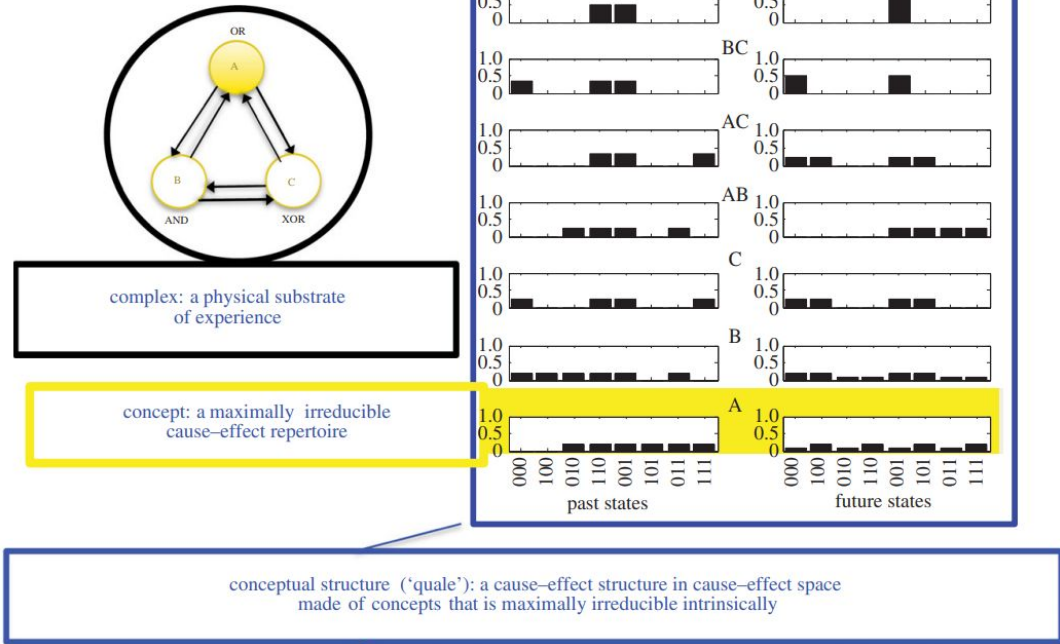


Abbildung: Giulio Tononi und Cristof Koch
Consciousness: here, there and everywhere?, Figure 4

Integrated Information Theory

- Ursache-Wirkung Raum hoch-dimensional (hier 2 x 8 Dimensionen)
- Jeder Stern markiert ein Konzept
- Position bestimmt durch Subsets
- Blaue Achsen, mögliche vergangene Zustände, grüne Achsen mögliche zukünftige Zustände
- Quale ist Konzeptionelle Struktur, die intrinsisch maximal unreduzierbar ist

conceptual structure ('quale'): a cause-effect structure in cause-effect space made of concepts that is maximally irreducible intrinsically

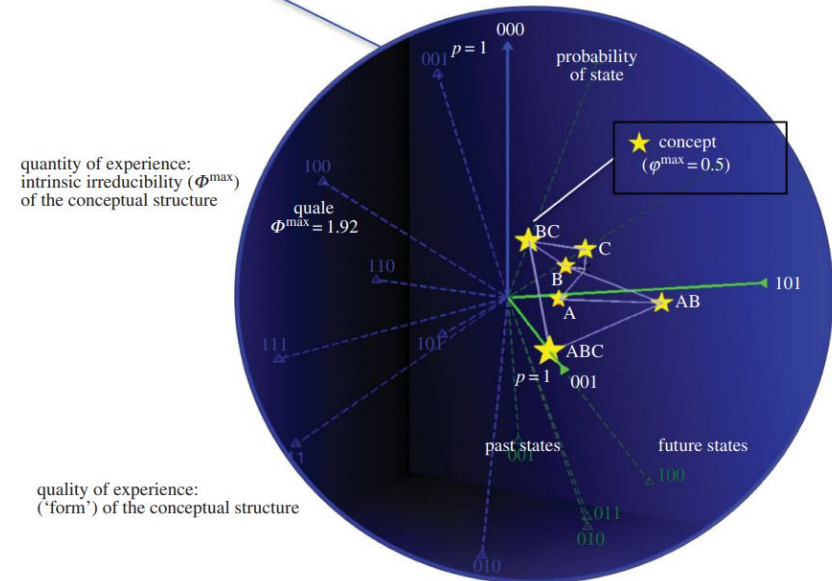


Abbildung: Giulio Tononi und Cristof Koch
Consciousness: here, there and everywhere?, Figure 4

Integrated Information Theory (IIT)

Folgerungen von IIT

- Axiome und Postulate treffen Aussage, dass Bewusstsein
 - Fundamental ist
 - Unabhängig vom Beobachter
 - Nur bestimmt durch Ursache-Wirkung Vermögen und Form der Möglichkeiten in Vergangenheit und Zukunft
- Qualität und Quantität von Erfahrung prinzipiell messbar
 - Qualität: intrinsische Unreduzierbarkeit der konzeptionellen Struktur
 - Quantität: Form der konzeptionellen Struktur
 - Beispiel zeigt, einfache Systeme sehr hoch-dimensional
- Jedoch, ähnliche Erfahrungen näher zueinander
- Mögliche Aufgabe von Psychophysik:
 - Zirkulare Natur des Farbraums auf Ursache-Wirkung Raum übertragen
 - Jedoch Form von Ursache-Wirkung Raum, ist Form innerhalb der Form der Erfahrung

Integrated Information Theory (IIT)

Folgerungen von IIT

- Bewusstsein kann sich entwickeln
- Bessere, untrennbare Vernetzung = höheres Bewusstsein
- Höheres Bewusstsein -> Bessere Anpassungsfähigkeit -> Bevorzugung durch Evolution
- Evolution durch natürliche Auslese führt zu Organismen mit hohem Bewusstsein
- Bewusstsein ist gestuft
 - Reichtum von Erfahrungen wahrscheinlich höher mit mehr und besser vernetzten Neuronen
- Entwicklung von Bewusstsein während Lebensspanne von Organismus
 - Säuglinge niedrigeres Bewusstsein als Erwachsene
 - Kann auch abnehmen, z.B. durch Drogen, Alkohol oder Krankheiten wie Demenz

Integrated Information Theory (IIT)

Folgerungen von IIT

- Während Phasen ohne nennenswertes Bewusstsein, z.B. tiefer, traumloser Schlaf
 - räumlich getrennte Komplexe haben wahrscheinlich $\Phi^{\max} > 0$
- Bewusstsein wäre dann nicht bei komplett 0, jedoch sehr gering verglichen mit alltäglichem
- IIT erlaubt mehrere Bewusstseine nebeneinander
 - Beispiel: Split-Brain Patienten, bei denen die Gehirnhälfte für Kommunikation keine Wahrnehmung für andere Gehirnhälfte besitzt
 - Wären zwei nicht verbundene Komplexe
 - Weitere “hoch-performante” Komplexe innerhalb des Gehirns möglich nach IIT, Messung prinzipiell durch integrierte Informations auf neuronaler Ebene

Integrated Information Theory (IIT)

Folgerungen von IIT

- Aggregate sind nicht bewusst
- Beispiel:
 - Satz mit 12 Wörtern, 12 Menschen. Jeder Mensch erhält ein Wort, kein Austausch. Kein Mensch kann auf Satz schließen
- Nur Maxima von integrierter Information existieren nach IIT
- Sprechen zwei Menschen miteinander: zwei Komplexe, hohe Werte von Φ^{\max}
- Sprechen zwei Menschen zusammen, formen sie ein System
 - Jedoch, es ist nicht maximal unreduzierbar, da beiden Komplexe mehr Information haben als das vereinte System
- Daraus folgt: kein gemeinsames Bewusstsein einer Gruppe, Landes, Team, etc.

Integrated Information Theory (IIT)

Folgerungen von IIT

- Komplizierte Systeme können bewusstlos sein
- Feed-forward Netzwerke sind unbewusst
 - Input hängt immer nur von externen Inputs ab
 - Output beeinflusst den Rest des Systems nicht
 - Beides Ausschlusskriterien für Komplexe
 - Netzwerk existiert nicht intrinsisch, also für sich selbst, sondern nur als “Zombie”
- Damit sind neuronale feed-forward Netze nicht bewusst
 - Auch dann, wenn input-output Verhalten von Menschen perfekt emuliert und der Turing-Test bestanden wird
- Feedback Netzwerke, also solche, bei denen die Inputs durch die Outputs beeinflusst werden, wären demnach bewusst

Integrated Information Theory (IIT)

Folgerungen von IIT

- Perfekte (Computer-)Simulationen des Gehirns?
- Nicht bewusst, denn Simulation des Gehirns virtuell
- Bewusstsein fundamentale Eigenschaft von physikalischen Systemen, mit realem Ursache-Wirkung Vermögen
 - System muss Vermögen haben, die Form des Raumes der vergangenen und zukünftigen Zustände zu verändern
- Computer, auf dem Simulation läuft möglicherweise bewusst, Bausteine des Computers müssten untersucht werden, nicht Software
 - Computer wahrscheinlich im Bereich von vielen kleinen Bewusstseine, die aber nicht zu einem großen geeint werden

Zusammenfassung und Diskussionsfragen

- Nach IIT ist Bewusstsein in Organismen und Systemen möglich
- Entscheidendes Kriterium ist, dass System für sich selbst existiert, Input nicht ausschließlich von außen kommt und Outputs Einfluss auf das System hat
- Was passiert mit meinem Bewusstsein unter Betäubung oder im traumlosen Tiefschlaf?
 - Kollabiert es?
 - Wiederherstellung durch Aufwachen oder Träumen?
 - Werden meine Erfahrungen durch temporäre Deaktivierung von Hirnregionen verändert?
- Was ist mit Computer-Simulationen?

Vielen Dank für die Aufmerksamkeit!



© ISTOCK / SVISIO (AUSSCHNITT), Quelle: www.spektrum.de

Quellen

- Tononi G, Koch C. 2015 “Consciousness: here, there and everywhere?” *Phil. Trans. R. Soc. B* **370**: 20140167: <https://royalsocietypublishing.org/doi/10.1098/rstb.2014.0167>
- Tononi, Giulio. “Consciousness as Integrated Information: A Provisional Manifesto.” *Biological Bulletin*, vol. 215, no. 3, 2008, pp. 216–242. *JSTOR*, www.jstor.org/stable/25470707