AI and Common Sense: Ambitions and Frictions

Martin W Bauer

Bernard Schiele

[Editors]

London

Routledge

Studies of Science, Technology and Society, Vol 58

ISBN: 978-1-032-62618-5 (hbk); ISBN: 978-1-032-62617-8 (pbk);

ISBN: 978-1-032-62619-2 (ebk); DOI: 10.4324/9781032626192

Chapter 1:

AI with Common Sense -

What Concept of Common Sense?

Martin W Bauer

[Oct 2023]

Common Sense is a theme that is present from the very beginning of AI; it was claimed already in 1958 that a particular computer program displayed 'common sense'. However, what is 'common sense' [CS], we must ask? This chapter seeks to clarify eight different concepts of 'common sense'. These historically derive from three historical types: Aristotle's 'koine aesthesis' of sensory integration, the 'natural reasoning' of the Scottish Enlightenment, and Vico's 'un-reflected moral community' which is both universal and culturally distinct. We also capture the tension of positioning CS in a vertical hierarchy or on a horizontal continuum of different forms of knowing. Being aware of different concepts of CS will enable us to assess critically claims made as to *'AI with CS*': which kind of CS is invoked?

The theme '*AI and Common Sense'* [CS] has been in evidence from the very beginning. The pioneer of AI, John McCarthy, wrote a paper (1958) famously entitled 'Programs with Common Sense'. Together with Marvin Minsky he had written a computer program called '*Advice Taker*' where they defined

> *'.. a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.' [78].*

Common Sense [CS] here means 'to draw inferences', i.e. to have the capacity to deduce conclusions from what is already known [memory stock] and from information given [input flow]. CS retains a constant presence in the AI debates ever since, it bothers the protagonists and constitutes a challenge for the AI field. CS creates frictions and is a source of resistance in this development. For some, who focus on developing expert systems, CS is irrelevant because experts are by definition beyond CS; for others, systems with CS are the last frontier to conquer and to complete our technological civilisation. However, we must rely on more competent commentators to tell this history of the encounter of AI and CS (e.g. Brachman & Levesque, 2022); some of it will surface in these chapters.

In the present chapter, I will point to several different ways in which we can understanding 'common sense' [CS]. By clarifying various uses of the notion of 'common sense', we seek to build a critical attitude and to ask the question: what do we possibly mean when the talk is of 'AI with CS'? We sharpen our understanding by being aware of different concepts as shown in table 1 below [CS-1 to CS-k] which are mobilised for or which are being left out from embodiment as AI. We will be able to clarify claims 'AI with CS', by pointing to different concepts of common sense that are invoked: 'AI with CS-1', or 'AI with CS-2', or 'AI has not/unlikely CS-k' etc. The 'Common Sense' of AI is thus better defined. While operationalising and objectifying our understanding of CS might bring clarity; this comes at the cost of **reification,** of limiting our understanding of CS by way of a technology standard. Thus, we seek to rescue the notion of 'common sense' from such a fate of reduction and the fallacy of nothing-but-ness: i.e. CS is nothing-but-what-'AI-with-CS'-defines-it-to-be.

In a first approach, CS compares unfavourably with expert, specialist, evidential and deductively derived knowledge.

Waldenfels (1982) reconstructs the genealogy of the perennial denigration of common sense in terms of the juxtapositions of Doxa versus Episteme, preliminary opinion against robust knowledge. In rehearsing this history of distinctions and cognitive hierarchies, CS falls at the bottom as mere Doxa. In the European context, this history of disqualification traces back to Greek philosophy that distinguishes appearances from reality, and sides mostly with Platonic against Sophist arguments. For the Sophists there is only Doxa as a matter of social conventions. While for Plato, Doxa is close to appearances that are not recognised as such, i.e. the perceptions of projected shadows that present themselves to those ignoranti who 'shackled in the cave' never had the privilege to leave the cave and to see the world in the true light of the sun. Doxa takes on the meaning of distorted cognition under poor conditions (see Heidegger, 2002; on the Plato's cave allegory in the 'Republic' and on Doxa in 'Theaetetus'). In modern parlance, this problem of demarcating a distorted Doxa from the dignified Episteme comes in three ways:

a) There is an **'epistemological cut or rupture'.** CS and science are very different; CS is deficient and misleading cognition. Science is counter-intuitive and avoids the pitfalls of natural language by applying rigorously quantitative concepts and methods. Science accumulates the evidence that arises from this effort which must ultimately replace CS to make the world a better place (e.g. Wolpert, 1992): where Doxa was there will be and must be Episteme.

b) Alternatively, the **continuity hypothesis** argues that science is little else than CS, only a bit more elaborate, refined and precise. Empiricism with its ethos of unprejudiced observation, experimental demonstrations, and a deep suspicion of rhetoric, tradition and authority, suggests that science is indeed 'purified' CS. Science basically is CS by another name.

There are variations on what 'purified' might mean here. Science seems to be an activity that is both rational-deductive and empirical-evidential; it guides human activity with

predictions and outlines options for choices. What distinguishes science from CS is that predictions and choices are made systematically, consciously and explicitly, and these are self-corrective through individual and collective learning. Thus, science is systematic CS, but as such nevertheless an extension of CS (e.g. Bronowski, 1951; Hoyningen-Huene, 2008).

In Luckmann's account (1987) science and CS are functionally equivalent for everyday live; we cook either with CS [fire, potatoes, vegetables, meat], or we cook scientifically [energy, carbohydrates, fibres, vitamins, proteins etc.]. However, historical modernisation reduced the reliance on CS and increases the reliance on science in all spheres of life. We read in the lifestyle magazine *'science tells which cat it good for you'* or *'Dutch randomised control trial: take cold showers in the morning to stay healthy'.*

c) Farr (1993), responding to the argument of 'unnatural science', argues for a third way, leaving it open whether there is a vertical gap between science and CS. The key is to **suspend judgement** and to study symmetrically both science and CS as exemplars of 'social representation' and to acquire the competence of 'cognitive polyphasia', i.e. being able to think with both (e.g. Jovchelovitch, 2008; Wagner & Hayes, 2005). The so-oriented social sciences map both CS and science and study their division of labour for clarity, not to eliminate either of them (Bangerter, 1995).

There are analogous ways of considering this symmetry: as 'belief systems' (Geertz, 1993), as 'language games and ways of life' (Wittgenstein, 1953) or 'life provinces' (Ryle, 1954) or in a post-Merton sociology of SSK (Lynch, 1993). Lynch famously does not find any differences in acting and account giving inside and outside the laboratory. When people put on the white coat and engage the lab benches, they do not go beyond the doings and thinking involved in any skilled activity. The claim to the contrary is at best a construction ex-post-factum to dignify the scientific life apart from everyday life. This boundary work accumulates favours and privileges, and science becomes a superior performance of knowledge.

Science is rooted in everyday concerns, what the phenomenological tradition calls the 'natural attitude'. Science builds from this attitude, and never transcends it. It is the phenomenological method of 'bracketing the natural attitude' [epoche] which allows us to recognise the communality between science and common sense, to remain neutral and to engage comparative analysis. Without the transcendental viewpoint from nowhere [i.e. to seek 'true illumination' outside Plato's cave], the epistemic contrast becomes co-production understood by a social psychology of assimilating and accommodating paradigms [mind sets, mentalities, thought styles] under conditions of inter-group competition (thought communities; see Fleck, (1979 [1935]; Kuhn (1962); Sammut & Bauer, 2021).

In a second approach, three historical types of common sense

Beside this long tradition of ordering vertically or horizontally, of denigrating 'common sense' and cognate concepts such as Doxa, everyday thinking, lay knowledge, and opinions, we must consider the parallel history of unfolding different concepts of 'common sense'. It appears that the current diversity of CS can be traced back to three historical origins:

- The 6th sense of Aristotle [384-32 BCA] as in 'koine aisthesis' [CS-1];
- The Scottish/English enlightenment (18th C) and its concern for everyday reasoning [CS-2, CS-3, CS-4];
- Giambattista Vico's (1668-1744) elaboration of the community of moral sensibilities, being universal and culturally distinct [CS-4, CS-5, CS-6].

Table 1: different kinds of 'common sense' to be explored

	Kinds of Common Sense
1	*CS-1 solves the 'binding problem', i.e. integration of multi-modal sensory experiences*
2	*CS-2 is a stock of universally recognisable knowledge to reason from*
3	*CS-3 is the quasi-rational judgment call; neither pure intuition nor formally rational*
4	*CS-4 marks the philosophy that in a moment of crisis stays clear of the errors of scepticism and dogmatism, by siding with lay people against elites*
5	*CS-5 is the communal sensitivity that binds us into joint attention and intentionality*
6	*CS-6 is what social psychology is all about [at least for some social psychologists]*
7	*CS-7 is what you can appeal to in order to bring fighting parties to they go off the rails in excessive polarization*
8	*CS-8 is what AI aims to simulate / therefore CS is what AI can do*
9	*?????? [still to recognise…..]*

A polemical impetus for both the Scots' and Vico's understanding of CS is to stay clear of Descartes' methodology (isolated individualism, mind-body dualism, axiomatic- geometric-deductive rationalism). Aristotle serves as point of return to our understanding of the integration of perceptual systems. These historical sources are further elaborated in modern parlance [CS-3, CS-4, CS-5, CS-6], and we can distinguish at least two uses of CS as 'empty signifiers' [CS-7, CS-8] that can be subversive in effect. All this makes no claim to be exhaustive, which is why in the summary table 1, a space is left open to add new kinds to this list.

Overall, we might say that the notion of CS requires two considerations: a) where to place CS in a vertical [hierarchy] or a horizontal [continuity] order of reasoning, and b) a variety of meanings of 'common sense' which derive from three historical types. Exploring eight different concepts of 'common sense' should enable us to critically examine any claim being made as to CS and AI:

… Upon coming across propositions like '*AI with CS' or 'AI is CS' ⇔ 'CS is AI' …*

We must ask immediately:

Which concepts of 'Common Sense' is invoked?

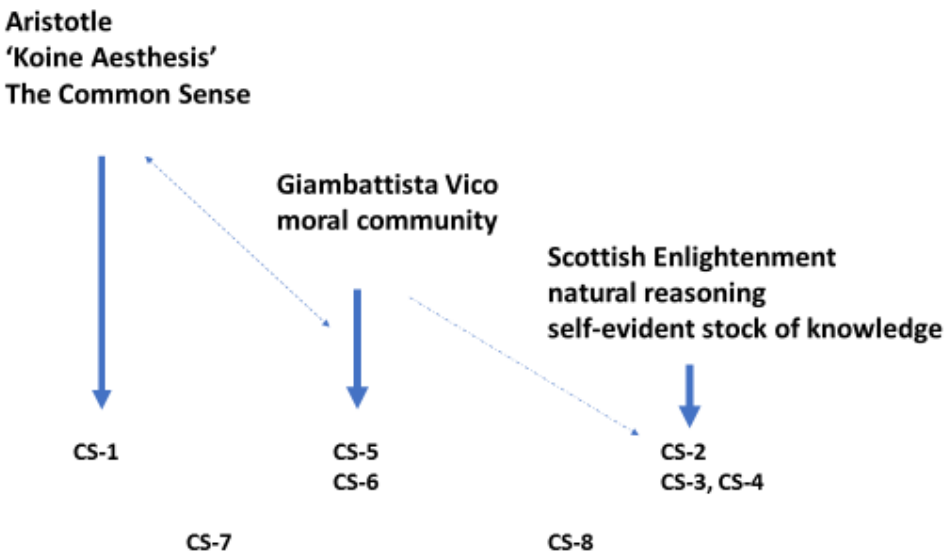

Figure 1: Three historical types of 'Common Sense' and some derivatives

Let us start by examining the three historical types: The Common Sense of Aristotle, the universal stock of knowledge to reason from of the Scottish Enlightenment, and the moral community with a history of Gambattisto Vico. These are traditions with overlaps and historical linkages.

Solving the 'binding problem' of sensory integration

CS-1 solves the 'binding problem', i.e. integration of multi-modal sensory experiences

The first is the oldest meaning, it goes back to the Greek term 'koine aisthesis' [common + sensation], better translated as 'common sensitivity'. In several essays, Aristotle [4th C BCA, axial times] refers to the 'sixth sense', the capacity that brings the five external senses into concert with multi-sensory features such as roughness, shape or magnitude. The senses are separable only in operational account; they are inseparable in perceptual power. First the heart, later the brain, were seen as the organ of this integration (Bennett & Hacker, 2003; p16). However, there remain textual difficulties and controversies in reconstructing the key functions marked by his term (see Gregoric, 2007; Hamlyn, 1968).

While the idea of separating senses into interior [common sense, memory] and exterior [five senses] has long be abandoned, the 'cross-modal binding problem' of sensory integration persists. How do we come from a faint sound [miou, miou], a visual impression of an elegantly moving four-legged shape, and a haptic fluffy stroke to recognise a 'cat', even before we apply the English word 'cat': if it purrs like a cat, walks like a cat, feels like cat … it

must be cat. This remains a serious problem for anybody who builds robots with sensory capacity. The transition from a pattern of sensations to perceiving tokens of kinds remains a philosophical puzzle (Siegel, 2012).

This idea of *The CS*, with a definite article, includes 'inner perception' of *'seeing that we see and hear'*, the concomitant monitoring, but not observation, of perceptual activity and thus the retrospective accessibility of mental states as 2nd order consciousness. Aristotle's psychology of the living organism postulates no subliminal, unconscious mental processes [1st order activity without 2nd order awareness], but knows of awareness of absent or failed perceptions [2nd order awareness without 1st order activity]. Aristotle thus re-enters modern discussions of consciousness and neural integration (Gregoric, 2007, 209ff). A popular version of this idea reappears with the higher cognitive function CS that integrates different forms of intelligence into practical, interpersonal actions that are intuitive, rapid and surprisingly effective (Gardner, 1983; 288f).

Self-evident knowledge and making good use of it

CS-2 is a stock of universally recognisable knowledge to reason from

A widely used concept of CS (no definite article) refers to everyday reasoning and available knowledge, the stock and the flow of inputs that are processed. This stock of knowledge is self-evident, no need for demonstration logical or otherwise, and widely shared and accessible. And to process these items, no expertise beyond being born and primary socialisation is needed because it is part of the human experience anywhere and everywhere. This 'mental dictionary' is not bound by any specific language but finds expression in the vernacular. In China, until recently, the school curriculum taught 'Changshi' [modern 常识 or traditional 常識]: what every young Chinese should know.

This CS is what the Scottish Enlightenment (e.g. Thomas Reid, 1710-1796) intended and what created the tradition of Common Sense philosophy (GE Moore, 1873-1958): providing a list of pre-reflective beliefs that are beyond doubt; these items don't need to be justified. The ambition is to be able to show that ways of thinking which are inconsistent with these basics are most likely in error (Boulter, 2007). Reid defined '*there are certain principles .. which the constitution of our nature leads us to believe, and which we are under a necessity to take for granted in the common concerns of life, without being able to give reasons for them … the principle of Common Sense'* (cited from Rosenfeld, 2011; 72). The polemical purpose is to trust our senses against sceptical detractors (Hume) and without the need for axiomatic rationalism (Descartes) because our senses do betray us only some of the time. A paradox might arise: in listing these items they are no longer taken-for-granted, because their self-evidence is subject to scrutiny: is the list incomplete or overextended?

Here is where AI research seems mostly situated, in attempts to upload **background knowledge** into the machine. The large scale project *Cyc Common Knowledge Base* [from EnCYClopedia] seeks to capture items in a formal representation across domains and in a cumulative manner since 1985 including ontologies-entities, types, concepts and assertions

(Davis & Marcus, 2015). Other forms of AI consider the entire internet as a freely available text base to teach a machine to parrot text production from large language models [LLMs] such as ChatGPT (Bender et al. 2021). OECD offers a classification of AI systems (OECD, 2022) to map the benefits and regulatory challenges of these new developments.

This kind of CS often carries a connotation of **acting below capacity** in the mode of a 'cognitive miser'. Many languages offer two words for CS to express ambivalence. In English the stress might be on 'common', in the elitist sense of 'pertaining to vulgar people' [House of Commons] in contrast to aristocratic [House of Lords]. In French, dubious and wide-spread 'le sens commun' contrasts to the more dignified but uncommon 'le bon sens'. Other languages make similar distinctions capturing the hierarchy of elite versus the people.

The community spirit that provides moral and political guidance

CS-5 is the communal sensitivity that binds us into joint attention and intentionality

CS so far is mainly an individual capacity, though widely distributed. However, CS is also a social phenomenon; the sense of community in mutuality that makes social life possible in the first place. Here the stress is on **CS for social integration**, the human capacity to orient toward others on mutual common good; i.e. the love and reality of community as a moral programme.

Giambattista Vico [1688-1744] elaborates this concept from renewing rhetoric as a virtuous commitment to truth and eloquence (Pompa, 1975, 27ff). CS is a condition of possibilities that are both universal and historically variable. Universals arise from solving three existential problems: authority (religion), sex and birth (kinship) and death (funerals, afterlife). Historical CS comes and goes with shifting institutions that in order to solve these tasks condition our beliefs, attitudes and actions. The 'moral community' is not the outcome of abstract reasoning, but of joint attention and joint intentionality on common problems. CS offers 'judgement without reflection', post-perceptual but pre-reflective reasoning (Schaeffer, 2004) that binds people together in rituals, conventions, and traditions.

The old rhetorical concept here is Doxa: widespread belief, opinion or common ground that can be invoked. No need to spell it out because taken-for-granted, but it resonates the arguments to persuasive effect. Doxa are *'those opinions ….. which are accepted by everyone or by the majority or by the wise … or by the most notable and reputable of them'* (Aristotle, in Barnes, 1984, 167). This CS is bound by the community of speakers and listeners who gather in conversation and dialogue. Classical rhetoric constructs here lists of 'common places' [topoi], generic ones that apply always, domains specific ones, or historically variable frames of public opinion.

Historically this CS is vested in religion (re-ligio, binding together), where members cultivate a sense of belonging and identity, and take council on how to behave. There seems to be a historical line from the Greek Polis, to later Church Councils, to church halls, to modern town hall meetings, theorised as 'third places', neither work nor family, where people deliberate, relate to authorities and reach a common understanding (Gadamer, 1960).

CS becomes reflexive as historical progress of society towards **communicative rationality,** in speech acts of deliberation, reasoning and collective sense making. Habermas (1984 and 2019) elaborates these acts oriented towards a common understanding and examines its historical conditions of possibility, including religious rituals and European medieval conflicts between knowing and believing. What makes this CS reasoning possible is (a) a public sphere that is inclusive among equals, non-violent, self-conscious and argument focussed, and that (b) involves participants who are committed to account for validity claims: [i] true about the world, [ii] truthful (sincere) about self and [iii] right in relation to others. This procedural 'sensus communis' guarantees dialogical reason over monological rationality that only oriented towards ego-centric success. By guaranteeing ritual and attitude, quality decisions can be achieved that are more likely to stick (legitimacy). Thus, in CS facts and values mix and cultivate a counter-factual aspiration, leaving us with a sense that something might be missing.

This procedural CS is further made concrete in 'deliberative opinion polling' (Fishkin, 2011). A cross-section of people is invited to examine and deliberate on controversial issues such as 'genetically modified crops in the food chain', 'genetic screening during pregnancy', or 'AI for medical diagnosis'. The outcome will be informed public opinion, which is an improvement on randomised polling. It makes for better decisions and adds legitimacy to responsive government.

Let us now look at some derivatives from these three historical types: quasi-rational thinking (CS-3), an expression of epistemic crisis (CS-4), and the object matter of social psychology (CS-6). We will finally examine two usages of CS as an empty signifier to launch appeals [CS-7 and CS-8] when 'times are out of joint in the state of Denmark'.

A quasi-rational thinking

CS-3 is the quasi-rational judgment call; neither pure intuition nor formally rational

Common is a process of casting judgement that falls between intuition on the one hand and formal-deductive rationality on the other (Hammond, 1996). CS offers a middle-third way in a false dilemma: no need to choose between intuition and rationality. The appeal to CS pulls an intuitive decision towards the more explicit pole and a formal judgement is freed to be more intuitive. It is the last stand against the Skylla of 'iron cage' expertise and the Charybdis of 'mad' intuition, seeking to avoid the pitfalls on either side.

CS recognises three obstacles of formal rationality, therefore not universally recommended. First, formal-analytical thinking needs a model, e.g. MAUD, multi-attribute-utility-decisions. If multiple models are available to structure a problem, the choice between such models is not again a formal-analytical choice, but one of 'satisfying common sense'. Models apply conditionally; at some point one will want to say 'this is how far the model gets me, now I have to make up my mind' [ibidem, 155]. Secondly, a formal model might be ready, but data of poor quality or hard to come by. E.g. forecasting models can predict the future, but lack of data is frustrating the effort; it would be nice to model Kondratjev long waves [40-50

years] of economic development with mathematical rigour for purpose of forecasting, but the fact of only five cycles since 1750 is insufficient data to fit any such model. Finally, there are situations for which no model is specified; these are called ill or unstructured problems, e.g. pure uncertainty over future trends.

Furthermore, formal-statistical models are only accessible for trained experts who know how to handle them. A theory of expert reasoning excludes the majority of stake holding humans from participating; at worst, experts disenfranchise lay people from making decisions justified by a technocratic mentality.

By contrast, intuition is universal and always at hand. Under time pressure, in confusing circumstances and with information overload when formal models are not readily applied, intuition is indicated. Time and resources permitting, intuition is however often displaced by questions such as 'why', 'how do you know' or 'what makes you say that'. Answers to deflect these challenges often resort of anatomical metaphors: my heart tells me, it is my gut feeling …. etc. But when pressed, intuitions might not carry the argument among rational people. However, as formalisms equally are limited [no conditions, no data, no model], the middle way of CS is called for to cast a judgement [ibidem, 158].

Then there is the dilemma between robustness and precision or between reliability and validity in making decisions which supports CS. Intuition brings robustness within large variation [being on target with a wide scatter], analysis bring precision but can be massively in error [being off-target with small variation, biased]. Intuition can be valid with low reliability; while formal analysis is highly reliability with no validity. This is well known in text-analytical methodology: interpretative methods make a lot of sense, but suffer from inter-reader reliability; semi-automatic text coding is highly reliable, but can be massively biased and hard to make sense of. What happens, now that automatic text classification becomes the procedural norm, reliability is conflated with validity and the dilemma is formalised out of sight and out of mind.

CS is thus the way to decide under pressures of time and resources and under uncertainty. Quasi-rationality opens up a continuum of 'imperfect reasoning' or bounded rationality [with a history of psychologists' names such as Brunswick, Heider, Simon, Gigerenzer, Kruglanski]. The compromise of 'supported intuition' is recommended, when not already the dominant mode of thinking in many walks of life, also for policy making and court rooms where civil servants strike a balance between rule following and specific contexts to achieve reason and justice [ibidem, p176]. This middle way tradition of a compromise is polemically juxtaposed against a paradigm that tests real-life decisions against formal models and typifies the deviances as biases or heuristics of 'irrationality' [psycho-physics of utility, slows and fast processing]. In applied economics, this becomes the stuff of Nobel Prizes (R Kahneman & A Tversky, Nobel Prize 2002; RH Thaler, Nobel Prize 2017). Predicting 'how CS sense gets it all wrong' can make you famous and can make you rich by exploiting human weaknesses in the choice-preserving-decision-architectures of 'nudging' [getting you to do something] and 'booming' [keeping you on track] on the cheap, because you can avoid expensive incentives on the one hand and costly regulations on the other.

An index of epistemic crisis and of a specific historical period

CS-4 marks the philosophy that in a moment of crisis stays clear of the errors of scepticism and dogmatism, by siding with lay people against elites

CS often refers to a particular period of European philosophy, mainly British enlightenment [18th C] associated with writers such as Thomas Reid [1710-1796] in Scotland and the Earl of Shaftesbury [1671-1713] in England (see Rosenfeld, 2011). Motivated by moral concerns, they elaborated an epistemological position that steers a middle way between scepticism - nothing can be known with certainty [Hume] - and the dogmatic rationalism [Descartes] built from one initial certainty 'I think, therefore I am' [cogito ergo sum], because our senses cannot be trusted. This CS philosophy found reception into German Enlightenment (Kuehn, 1980; Gadamer, 1963), and remains influential in modern philosophy of natural language (Ryle, 1954) and in the 'English character': one of gentle common sense against highflying distractions from the continent (Burke & Pallares-Burke, 2016).

CS in this sense might have little significance beyond elucidating an episode of conflict and strive in the history of philosophy. However, based on this episode, references to CS can henceforth be read more generally as expressions of an epistemic crisis, when people are disoriented by authoritarian coercion on the one hand and excessive solidarity and tribal identity politics on the other (Lindenberg, 1987). People refer to CS when something important seems already to be lost. Modern science, a policy authority, displaces CS and aggravates this problem by removing the ground from which to launch this appeal (Luebbe, 1987).

Though recent reading has discovered here the sources of a 'critical psychology' for the 21st century (Billig, 2008). By harking back to this traditions of reasoning, it reconstructs the roots of a psychology that is a-priori social in nature, avoids mentalism as strongly as behaviourism, and recognises the significance of language and rhetoric in the formation of attitudes in cognition, judgement and motivation.

The 'dynamic object' of a truly social psychology

CS-6 is what social psychology is all about [at least for some psychologists]

CS is the very object matter of a field of psychology. For Smedlund's programme of critical research (1997), it is the task of social psychology to clarify the concepts of human behaviour that are already embedded in ordinary language before any scientific progress can be achieved. Psycho-logical Common Sense is embedded in vernacular languages. What social psychology often proposes as 'theory' is thus revealed as 'platitudes' of a language that are otherwise invisible. This is exemplified with Bandura's very successful theory of self-efficacy (Smedslund, 1978); what are claimed to be empirical observations are the psycho-logically implications of assumptions ready-made in ordinary English language. The resonance of a theory in this common sense might even be an explanation for its success.

Correia Jesuino (2011) argues for a societal psychology with focus on common sense. This is the mark of the European tradition of a societal psychology, against the overgeneralisation of an US dominated model of social psychology. CS is the very object matter of societal psychology; experience and behaviour is understood within the 'social representation' framework, i.e. the acts and values which are identified and encouraged within the taken-for-granted matrix of a social group. This manifests itself in inter-group conflicts based on hetero- and auto-stereotyping, with confirmation biases in the information processing: East against West, North against South, nation against nation, milieu on milieu, tribe on tribe, and corporations and their competitors.

These structures of experiences, sense-making and explanations are historically dynamic and develop through social influence. When challenged by deviance, dissidents or newcomers, CS can accommodate what is unthinkable, make is possible, then probably, and end up with a new taken-for-granted. Continuous updating of CS occurs in cycles of normalising the unusual, assimilating the unfamiliar, and accommodating the challenges in considered adaptations. Modalities of these social influences including the power of crowds, leaders, conforming to peer pressure, conversion, obedience to authority, compliance with norms, persuasion, and resistance against the normative power of the fait-accompli or designed artefacts. For each inter-subjective influence, we might examine inter-objective extensions, as when we fix peer pressures into legal sanctions, or when we replace the authority of a policeman with a speed bump. Sammut & Bauer (2021) have recently provided an overview statement, culminating in a *periodic table of social influences,* of ways of updating CS through cycles of normalisation, assimilation and accommodation.

A political appeal to bridge polarized extremes

CS-7 is what you can appeal to in order to bring fighting parties to their senses when they go off the rails in excessive polarization

The appeal to CS is used in everyday speech to subvert some obvious stupidity, absurdity or illogicality. By appealing to CS, we are able to identify states of affair that cry out to be called stupid, irrational, unreasonable, or non-sensical. Famously, the 20th century was once characterised as the century of the 'common non-sense' (Chesterton), thus denouncing both Fascism and Communism as stupid ideologies.

One can thereby appeal to others to *'be reasonable', or 'let us be reasonable'* …. And assumes that everyone knows what the standards for this are. Or we can refer to CS by way of a rhetorical question, *'Is it not CS, that Google should pay tax in the country ….?'* [So in the British parliament in a debate on tax avoidance]. This appeals to social interaction among equals (Lindenberg, 1987, 203), and reveals a situation that, as with the legal avoidance of tax, is out of sync with common sense: *if you do business in this country, have such a presence and a large customers base, it must be clearly wrong, if your entity does not pay taxes in this place?* No further proof is needed, as the question is actually rhetorical to which the answer is given by common sense unspoken.

If this appeal is no longer possible, fundamentals have already moved too far. The common sense has been silenced by authoritarian language policing or by overwhelming tribal solidarity of us-against-them identities (Lindenberg, 1987). Where common sense truth can no longer be spoken, the polarised tribalism only sees enemies rather than legitimate counter-arguments arising from common ground.

An operational definition inspired by technical capabilities

CS-8 is what AI aims to simulate / therefore CS is what AI can do

Finally, we consider a meaning of CS that is both very technical and the very focus of polemical commentary. Brachmann & Levesque (2022) write about 'AI with CS'; they defend the paradigm of computing with symbolic representation to simulate 'human reasoning' in explicit models. They argue that this actually allows to simulate common sense reasoning. They contrast this with the current models of 'deep learning' AI where the system is a black box learning to efficiently detect hidden cues that discriminate for purposes of classification and predictions. In the end, nobody knows on which cues the classification was achieved; but the result can be tested for effectiveness of prediction. For example, in social media advertising [so my students tell me], females receive pregnancy related product advertising on the basis of their internet search history before they consciously think of becoming pregnant, known famously as 'the machine knows you better than you know yourself'. One could argue that Brachmann & Levesque want to put CS into the machine to avoid this type of predictive manipulation pure and simple.

The nothing-butness fallacy in AI equates what 'is' [intelligence exists, being] to 'what can be made' [machines intelligence, design], and confidently declares that this is all there is; if simulation reveals the process, forget all else. However, this ultimately conflates the model with the original, and commits a kind of secular 'iconodulance', i.e. violating the biblical injunction against making images of important matters and risking reducing matters to technical designs. In the same way as the priest worries if the believer takes the icon to be God, the secular observer worries about confusing the model with reality. A recent version of this confusion is to declare that, because ChatGPT produces plausible arguments on the basis of stochastic models of language, arguing people are just 'stochastic parrots' [like the machine], thus confusing people with the model (see critically Weil, 2023; Bender et al, 2021).

A recent paper in the Bulletin of the Atomic Scientist took aim at the existential threat that AI will in the near future eclipse humans on a whole range of tasks [GAI with singularity, super intelligence]. We are reassured that this eclipse is unlikely to occur. However, we should be anxious about how AI operations, however defined, become the standard against which humans are assessed and then humans come to be seen 'deficient'. Thus reality is assessed on the model and not the model on reality. We are not assessing robots against humans, but humans against robots as effective optimising machines. This reversal of judgement is variously dangerous. What is required is AI to human measure, i.e. human-

centred, human compatible designs, not 'iconodulant' aspirations to superhuman enhancement (Russell, 2019).

This underlying fallacy, known in psychology and physics as operationalism – temperature is what the thermometer measures, or intelligence is what the intelligence test measures - conflates the measure with the phenomenon, the measurement model with reality, the sign with the referent, or the image with the original, and ultimately turns the icon into God. The computer, initially a useful metaphor becomes 'the mind'; and in the end, by way of talking the mind is nothing-but-a-computer against which we benchmark actual human minds. The partial and particular aspect of any model metaphor is thus lost in translation: the circumspect proposition [X is Y / under the aspect A] before long is simplified; maybe for reasons of creating a meme on social media, it is reduced to [X is Y] in common parlance. The all-important proviso 'under the aspect of A' is dropped. It is a wide-spread pathology, even of scientific discourse, to conflate the impressive numerical model, whether statistical, analogue or digital, with the actual reality. Models seem to have persuasive power.

Conclusion – beware of conflating the model with reality, the signifier with the signified

This fallacy of taking the sign for the referent triggers historically the alarm in monotheistic religions worried about fetishism and idolatry, i.e. the breaking of the 2nd Mosaic Law which forbids the drafting and venerating of images, and thereby risking to mix it up with the real thing; it derives from the 1st Law about a jealous God that tolerates no others (Halbertal & Margalit, 1992) and cleaned out the pantheon. Such religions cultivate a sensitivity against the deification of artefacts as a betrayal of the one and only and warn against wild projections of superhuman omni-potency markers of self-deification. This temptation is checked by movements of recurrent iconoclasm. The high priests worrying that believers are taking the icon for God, call to smash the icons. This original impetus of iconoclasm is preserved in secularised versions; the religious sensitivity about 'things on the wrong track' is transferred to a secular suspicion of 'ideology and false consciousness', and against 'reifying concepts' in a rigid framework and at the same time losing sight of doing so (Berger & Luckman, 1966).

Earlier criticism of AI came up with the image of *'the drunkard looking for the lost key only under the lamp post where there is light'* [the streetlight effect]. This story tries to show that AI creates machine capacities within a theoretical scheme; it is operating under a frame that sheds the light. Taking any particular frame as exhaustive of human capacities is mistaking convenience and present technical prowess for the final direction, thus also cutting short the scientific ambitions of understanding. While there will be other and better frames, there will always be AI only within a modelling frame. Any application of that framing has unintended consequences that in fact need our attention beyond the good intentions of the designers (Dreyfus & Dreyfus, 1986). Also in the most recent renaissance of AI, designers might be good at making things, but often ignorant of what they are actually doing, which keeps observers busy assessing and monitoring the unanticipated consequences of designs once they are in-use (Edgerton, 2006; Merton, 1936).

We will keep discussing this problem of confusing models with reality as the continued challenge of Artificial Intelligence to and by Common Sense. The first question to ask is: *'AI with CS', which kind of CS is invoked here*? The second point to raise: *'AI beyond CS', even worse to contemplate.*

References

- Bangerter A (1995) Rethinking the relation between science and common sense: a comment on the current state of Social Representation theory, Papers on Social Representations, 4, 1, 2-18.
- Barnes J (1984) The complete works of Artistotle – the revised Oxford translation, Princeton, PUP, vol1 and vol2.
- Brachman RJ and HJ Levesque (2022) Machines like us – towards AI with common sense, Cambridge MA, MIT Press.
- Bender EM, A MCMillan-Major, T Gebru and S Shmitchell (2021) On the dangers of stochastic parrots: can language models be too big? FAccT '21, March 3–10, 2021, Virtual Event, Canada ACM ISBN 978-1-4503-8309-7/21/03. https://doi.org/10.1145/3442188.3445922
- Bennett MR and PMS Hacker (2003) Philosophical Foundations of Neuroscience, Oxford, Blackwell Publishers.
- Berger P and T Luckmann (1966) The Social Construction of Reality – a Treatise in the Sociology of Knowledge, London, Penguin.
- Burke P and MLG Pallares-Burke (2016) Os ingleses, Sao Paulo, editor context
- BILLIG, Michael. The Hidden Roots of Critical Psychology: Understanding the Impact of Locke, Shaftesbury and Reid. London, 2008.
- Boulter S (2007) The Rediscovery of Common Sense Philosophy, Basingstoke, Palgrave Macmillan.
- Correia Jesuino J (2011) Back to common sense, in: J Pires Valentim (ed) Societal approaches in social psychology, Bern, Peter Lang, pp35-60.
- Davis E and Marcus G (2015) Commonsense reasoning and commonsense knowledge in artificial intelligence. Communications of the ACM58, 9 (August 2015), 92-103. DOI:
- Dreyfus H and S Dreyfus (1986) why computers may never think like peope, Technology Review, January.
- Edgerton D (2006) The shock of the old – technology and global history since 1900, London, Profile books
- Farr R (1993) Common sense, science and social representations. Public Understanding of Science, 1993, 2, 111-122.
- Fishkin, JS (2011) When the People Speak: Deliberative Democracy and Public Consultation, Oxford : Oxford University
- Fleck L (1979) Genesis and development of a scientific fact, Chicago, CUP [German original, 1935]
- Gadamer HG (1960) Wahrheit und Methode, Mohr Verlag [truth and methods]
- Gardner H (1983) Frames of mind – The theory of multiple intelligences, London, Paladin Granada Publishers
- Geertz C (1993) Local knowledge [chapter 4; CS as a cultural system], London, Fontana Press.
- Gregoric P (2007) Aristotle on the Common Sense, Oxford University Scholarship

- Habermas J (1984) The theory of communicative action, 2 vols, Cambridge, Polity Press.
- Habermas J (2019) Auch eine Geschichte der Philosophie, 2 Vols, Frankfurt, Suhrkamp.
- Halbertal M and A Margalit (1992) Idolatry, Cambridge MA, Harvard University Press.
- Hamlyn DW (1968) KOINE AISTHESIS, The Monist, Vol. 52, No. 2 (APRIL), pp. 195-209
- Hammond KR (1996) Human judgement and Social Policy – irreducible uncertainty, inevitable error, unavoidable injustice, Oxford, OUP.
- Heidegger M (2002) The essence of truth – On Plato's cave allegory and Thaetetus, London, Bloomsbury [vom Wesen der Wahrheit, Klostermann, 1988]
- Hoyningen-Huene P (2008) Systematicity: the nature of science, Philosophia, 36, 167-180.
- Jovchelovitch S (2008) The rehabilitation of common sense: social representations, science and cognitive polyphasia, Journal for the Theory of Social Behaviour, 38, 4, 431-448.
- Kuehn M (1980) Scottish common sense in Germany 1768-1800, PhD dissertation, Montreal, McGill University.
- Kuhn TS (1962) The Structure of Scientific Revolutions, Chicago, CUP.
- Lindenberg S (1987) Common sense and social structure: a sociological view, in: Van Holthoon and D R Olson (eds) Common sense: the foundations for social science, Lantham, MD, University of America Press, p199-215.
- Luckmann T (1987) Some thoughts on common sense and science, in: Van Holthoon and D R Olson (eds) Common sense: the foundations for social science, Lantham, MD, University of America Press, p179-198.
- Luebbe H (1987) Die Wissenschaften und ihre kulturellen Folgen – Ueber die Zukunft des Common Sense, Rheinisch-Westfaelische Akademie der Wissenschaften, Vortraege G285, Opladen, Westdeutscher Verlag.
- Lynch M (1993) Scientific practice and ordinary action – Ethnomethodology and social studies of science, Cambridge, CUP.
- McCarthy J (1958) Programs with common sense, session 1 paper 3 [on-line source]
- Merton RK (1936) the unanticipated consequences of purposive action, American Sociological Review, 1, 6, 894-904.
- OECD (2022) The OECD frame for the classification of AI systems [www.oecd.ai/classification]
- Pompa L (1975) Vico – a study of the ‚New Science', Cambridge, CUP
- Rosenfeld S (2011) Common Sense. A political history, Cambridge, MA, Harvard University Press
- Ryle G (1954) The world of science and the everyday world, in Dilemmas. Cambridge, CUP, 68-81.
- Sammut G and MW Bauer (2021) The Psychology of Social Influence – Modes and Modalities of Shifting Common Sense, Cambridge, CUP
- Siegel S (2012) The contents of visual experience, Oxford, OUP
- Russell S (2019) Human compatible – AI and the problem of control, London, Penguin.

- Schaeffer JD (2004) Commonplaces: sensus communis, in: Jost, W and W Olmsted (eds) A companion to rhetoric and rhetorical criticism, Oxford, Blackwell, pp278-293.
- Smedslund J (1997) The structure of psychological common sense, Mahwah, LEA Publishers
- Smedslund J (1978) Bandura's theory of self-efficacy: a set of common sense theorems, Scandinavian Journal of Psychology, 19, 1-14.
- Wagner W and N Hayes (2005) Everyday discourse and common sense – the theory of social representations, Basingstoke, Palgrave Macmillan.
- Waldenfels B (1982) The Despised Doxa – Husserl and the continuing crisis of Western Reason, Research in Phenomenology, 12, 1, 21-38; (translated from Japanese SHISO in by J Claude Evans).
- Weil E (2023) You are not a parrot, NY Intelligencer [accessed 14 March 2023] https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html
- Wolpert, L. (1992). The unnatural nature of science, London, Fabor & Fabor