The looping effects of human kinds (Hacking 1996)
Published within *Causal Cognition: A multidisciplinary debate*. Part IV: Understanding social causality

Main Theme: "...how a causal understanding, if known by those who are understood, can change their character, can change the kind of people they are. That can lead to a change in the causal understanding itself."

Q1: What are Human Kinds?

- Kinds as systems of classification
    - "Kinds of behavior, act, or temperament…are human kinds if we take them to *characterize kinds of people*"
- Kinds that are studied in the "marginal, insecure, (oof) but enormously powerful human and social sciences"
- Kinds about which "we" would like to have:
    - Systemic knowledge
    - Classifications that can form general truths about people
    - Generalizations that are sufficiently strong and law-like about people, their actions, and their sentiments
- They should be causally relevant, i.e. we want to be able to predict what individuals will *do* or how they will respond to attempts to *modify* their behavior
- Lots of disclaimers about his take on natural kinds. Basically, he avoids commenting on what qualifies as a "natural kind" because that don't really matter to the distinction
    - What's keeping him from lumping human kinds in with natural kinds? Convenience, mainly.
- He restricts human kinds to those that are ("at least at first sight") *peculiar* to people in a social setting
    - Thus things like mass, longevity, etc., which are attributes of people (and other living things) regardless of where they are, are not counted as human kinds
    - Ok, but what if we view everything as a consequence of biology/human actions as consequences of neuroelectrical/biochemical/"natural" things? (e.g. an "alcoholic gene")
- Regardless of the above, he thinks that "there is little difficulty in picking out characteristic human kinds"
    - They are 1) relevant to some of us, 2) kinds that primarily sort people, their actions, and behavior, 3) are studied in the human and social sciences, *4) are projected to form the idea of a kind of person*
        - Example: idea of "the homosexual"
- "Human kinds usually present themselves as scientific and hence as *value free*, but they have often been brought into being by *judgements of good and evil*"
    - Example: study of suicide—from morally/religiously Bad to the result of extreme emotional distress/cry for help
- "Human kinds are of many categories"
    - Negative example: "women-fire-and-dangerous-things"

    - Positive examples: census categories, like race, gender, nationality, etc., *when endowed with their social connotations*
- What does he mean by "with their social connotations"? He explains this with the teenage pregnancy example
    - Teenage pregnancy has always been a thing, but only took on cultural meaning + relevance as a classification at a certain point in time: it is "completely grounded in nature, but is a human kind…only in a certain social context"
- Some other examples he brings up: adolescence, child abuse, multiple personality disorder
- "Human kinds begin in the hands of scientists of various stripes", but "people of the kind may rise up against the experts"
    - "We" would like to know about these kinds in order to help "them": the search for human kinds is intertwined with *prediction and reform*
- What comes first, classification or causal connections? He argues they come together

Q2: What's so special about Human Kinds?

- Back to the issue of distinguishing between human vs. natural kinds
    - Backdrop on the positivist and historicist view that all human kinds are natural kinds
    - Agrees with the historical take on natural kinds, but disagrees about the "end"—that the inclusion of human kinds with natural kinds is a step towards the causal understanding of *nature*
- What he is not arguing over: the *Verstehen* dispute and constructionist dispute
    - "We are stuck with human kinds that demand causal analysis rather than *Verstehen*"
    - "I operate as if there were no vital contradiction between realism and constructionism"
- A fundamental difference: **Looping**

Looping

- Human kinds are laden with values and have moral connotations
    - There are attempts to strip human kinds of values by biologizing them; this is done in the hope that causal connections between kinds are more intelligible if they operate at a biological level
- "Human kinds are kinds that people may or may not want to be"
- The act of classifying people changes them and can change their past
    - Causal relationships between kinds are changed, and even confirmed to the point of essential definitional connections (example: multiple personality disorder and repeated childhood trauma)
- Because the kind itself changes, there is new knowledge to be had about the kind
- "The greater the moral connotations of a human kind, the greater potential for looping effect"

“A Motley Collection” of More Kinds

- “In my opinion there are many more types of human kinds…but I do not think there is any core (‘prototypical’ kinds)”
- **Second-order kinds**
    - Normalcy–nothing is “just” normal
    - “Abnormal” psychology
    - Any human kind explained in terms of deviation from the normal is partly descriptive and partly evaluative
- **Biologized kinds**
    - Biologizing human kinds does not make them immune to looping effects
    - Examples: alcoholism—scientific vs moralistic view
    - The “wandering” of kinds—how a human kind reacts to the way in which the people who fall under the kind themselves react to being treated in the way science dictate
- **Inaccessible kinds**
    - But what if the people of a kind are not able to realize they are part of a classification? Like babies
        - The looping involves a larger human unit, like a family
    - Example: autism
- **Administrative kinds**
- **Self-ascriptive kinds**
    - Two axes for human kinds: one of the natural sciences, the other bureaucratic-statistical
        - A third: individualism
    - Self-identification; a claiming of rights to their own knowledge

Basic Objects in Natural Categories

ELEANOR ROSCH, CAROLYN B. MERVIS, WAYNE D. GRAY, DAVID M. JOHNSON, AND PENNY BOYES-BRAEM

Aim: To prove that the world is structured and contains "intrinsically separate things".

structured: *if the world formed a set of stimuli, not all possible combinations occur with equal probability (eg.: an object with the visual appearance of a chair is more likely to have "sit-on-ableness" than objects with the appearance of a cat.)*

Motivation: Argument against Leach 1964: " … *the physical and social environment of a young child is perceived as a continuum. It does not contain any intrinsically separate 'things.' The child, in due course, is taught to impose upon this environment a kind of discriminating grid which serves to distinguish the world as being composed of a large number of separate things, each labeled with a name*"

Scope: concrete objects; taxonomies defined in English; (one experiment in ASL)

Notes:

- Some pairs, triples, or n-tuples of attributes are quite probable leading to the formation of categories.

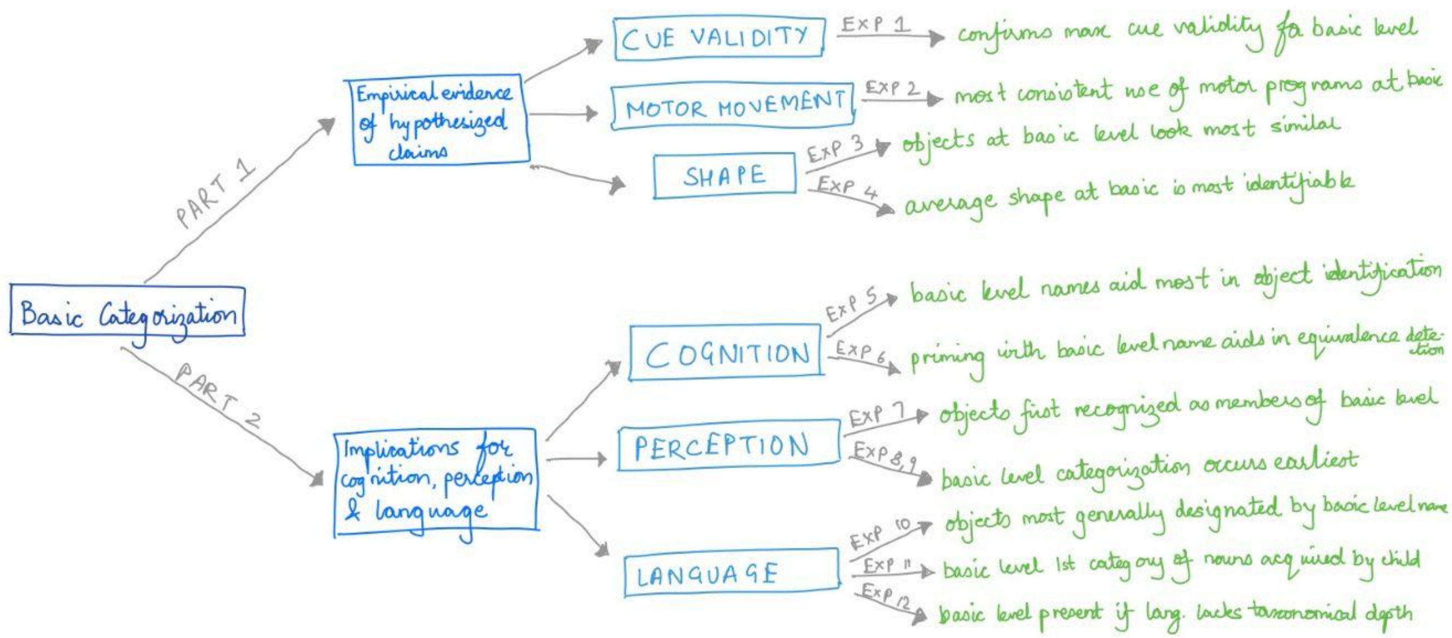
  Category : *objects which are considered equivalent (for a particular combination of attributes) e.g. dog, animal.*

  Taxonomy : *system by which categories are related to another by means of class inclusion. The greater the inclusiveness of a category within a taxonomy, the higher the level of abstraction (example: vehicle v/s car)"*

- Taxonomies are man-made linguistic categorizations of real-world concrete objects. The author argues that within them, there is one level of abstraction at which the most basic category cuts can be made (this "basic" level is what the author believes to be reflective of natural discontinuities in the real world, which she discusses and sets out to prove in the rest of the paper).

- Basic categories can be defined in terms of : a) cognitive economy and b) cue validity
  - Cognitive economy : On one hand, we want to know of as many attributes as possible using one label (which would lead to an infinite number of labels / categories); on the other, we want to reduce the infinite differences among stimuli to cognitively usable proportions. The basic level is a combination of the two: most *general* and *inclusive* level at which categories can delineate real-world correlational structures.
  - Cue Validity : "probabilistic concept"; the validity of cue x as a predictor of a category y (conditional probability of y/x) increases as the *frequency(x, y)* increases and decreases as the *frequency(x, ~y)* increases.
    - The cue validity of an entire category is the summation of the cue validities of each of the attributes of the category.
    - The author argues that total cue validities are **maximized** at the basic level. Categories one level more abstract will be superordinate categories (e.g., furniture, vehicle) whose members share only a few attributes among each other.
    - Categories below are subordinate categories (e.g. kitchen chair, sports car) wherein many attributes overlap with other categories (for example, kitchen chair shares most of its attributes with other kinds of chairs, which means "sit-on-ableness" cannot be used to differentiate between kinds of chairs).

The rest of the paper has two parts. **Part 1** comprises four experiments aimed at providing evidence supporting the concept of basic level objects (as the author defines it to be above). **Part 2** studies the implications of this concept for a) cognitive representation of categories; b) perception of objects; c) language acquisition and evolution; through 8 experiments (all experiments conducted on real and not artificial NNs, for a change :))



Part 1: Empirical evidence for hypothesized claims

Experiment 1:

Hypothesis: The basic level of abstraction in taxonomies of common objects, is the most inclusive level at which objects of a category possess common attributes.

Method: The participants listed attributes for all categories (as shown in table) 1) from memory; 2) by observing physically present objects. Finally, the number of common attributes for each category was computed as shown below. The category at which the maximum delta occurs, can be thought of as the categorization which provides for maximum information.

Superordinate	Basic level	Subordinates	
		Nonbiological taxonomies	
Musical instrument	Guitar	Folk guitar	Classical guitar
	Piano	Grand piano	Upright piano
	Drum	Kettle drum	Base drum
Fruit[a]	Apple	Delicious apple	Mackintosh apple
	Peach	Freestone peach	Cling peach
	Grapes	Concord grapes	Green seedless grapes
Tool	Hammer	Ball-peen hammer	Claw hammer
	Saw	Hack hand saw	Cross-cutting hand saw
	Screwdriver	Phillips screwdriver	Regular screwdriver
Clothing	Pants	Levis	Double knit pants
	Socks	Knee socks	Ankle socks
	Shirt	Dress shirt	Knit shirt
Furniture	Table	Kitchen table	Dining room table
	Lamp	Floor lamp	Desk lamp
	Chair	Kitchen chair	Living room chair
Vehicle	Car	Sports car	Four door sedan car
	Bus	City bus	Cross country bus
	Truck	Pick up truck	Tractor-trailer truck
		Biological taxonomies	
Tree	Maple	Silver maple	Sugar maple
	Birch	River birch	White birch
	Oak	White oak	Red oak
Fish	Bass	Sea bass	Striped bass
	Trout	Rainbow trout	Steelhead trout
	Salmon	Blueback salmon	Chinook salmon
Bird	Cardinal	Easter cardinal	Grey tailed cardinal
	Eagle	Bald eagle	Golden eagle
	Sparrow	Song sparrow	Field sparrow

Note: Lower levels are assumed to include all attributes listed at higher levels; only attributes new to a lower level are listed.

Tool	*Clothing*	*Furniture*	*Bird*
make things	you wear it	no attributes	feathers
fix things	keeps you warm	*Chair*	wings
metal	*Pants*	legs	beak
Saw	legs	seat	legs
handle	buttons	back	feet
teeth	belt loops	arms	eyes

(continued)

<u>*Results / Discussion:*</u> Hypothesis significantly supported. For biological taxonomies, the basic level appeared to be the next higher level in the taxonomy than had been proposed by anthropological evidence. Results of attributes listed from memory do not differ from when attributes are listed by observing physically present objects —> findings are not an artifact of the way in which object names are stored in memory.

However, they are an artifact of the knowledge / expertise of the observer.

Experiment 2:

<u>*Hypothesis:*</u> Basic level objects are the most inclusive categories for which highly similar sequences of motor movements are made to objects of the class.

<u>*Method:*</u> Obtain subjects' descriptions of the body and muscle movements they make when interacting with objects and tally the number of motor movements in common for the category terms at various levels of abstraction. As a

check on the validity of their descriptions from memory, models performed the main general activity which had been elicited by the object name with a sample of actual objects (e.g., a model sat down in a chair), and subjects described the actual body and muscle movements which they observed.

Tool	*Clothing*	*Furniture*	*Bird* (look at)
Hand: grasp	Eyes: scan	Eyes: scan	Eyes: scan
Fingers: grasp	Hand: grasp	*Chair* (sit on)	pursue
Hammer	*Pants* (put on)	Head: turn	look up
Arm: extend	Hands: grasp	Body: turn	squint
Hand: big grasp	Arms: extend	move	blink
position	Back: bend	back	Head: turn
Fingers:	Feet: position	position	pursue
position	Knee: bend	Knees: bend	Neck: tip back
Other hand:	Leg: raise	Arm: extend-	
position	extend	touch	*Sparrow*
Body: bend	Foot: raise	Waist: bend	Eyes: scan
Neck: bend	extend	Butt: touch	pursue

(continued)

Results / Discussion: Hypothesis supported, both for number of common movements and total number of movements;

- Superordinate: few, if any, motor movements in common.
- Basic level : many specific movements made to all members; similar movements described by a sufficient number of subjects to form a general picture of movement sequences for that category.
- Subordinate: do not differ significantly from basic

Experiment 3:

Hypothesis: Basic level is the category in which the shapes of objects would show a gain in objective similarity over the next higher level of abstraction.

Method: Random pictures for each category selected, normalized, adjusted for orientation etc. Outlines of the shape of each object were obtained using these pictures, as shown below. Ratios of overlaps of normalized shapes were computed for calculating similarity [$= A \cap B / (A \cup B - A \cap B)$]

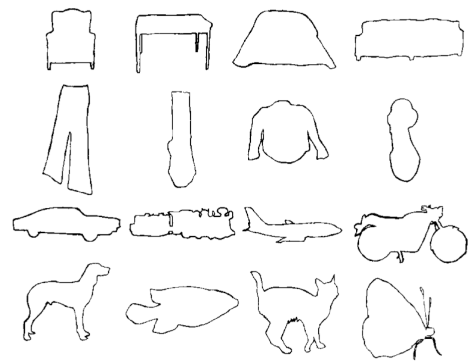


Fig. 1. Examples of traced outlines of pictures used in Experiment 3. One example of each basic level object from each superordinate category is shown.

Results / Discussion: Hypothesis is supported. A large and consistent increase in similarity of the overall look of objects (as measured by in- crease in the ratio of area of overlap to nonoverlap of normalized shapes of the objects) was obtained for basic level over superordinate categories. A significant but significantly smaller increase in similarity was obtained for subordinates over basic level categories.

Experiment 4:

Hypothesis: Basic level is the category in which shapes are sufficiently similar to render the shape of an average of more than one member of the category identifiable as a category member.

Method: Same outlines chosen as experiment 3. Outlines from objects for each category were averaged, and subjects were asked to label which category the average belongs to.

Results / Discussion: Hypothesis supported, averages of superordinate objects could not be identified as such better than chance; basic level objects were the most inclusive categories at which objects were readily identified. Furthermore, sub- ordinate object averages were no more identifiable than were the basic level averages.

Part 2: Implications for cognitive, perceptual and linguistic development

Experiments 5 and 6: Basic object categories are the most inclusive categories which can be represented by a code concrete enough to be called an image. It's the level at which one has a mental model / image of the class

Experiment 5

Hypothesis: Basic object categories are the most inclusive categories which can be represented by a code concrete enough to be called an image. It's the level at which one has a mental model / image of the class

Method: Pictures of objects of all categories viewed through a mask constructed of random shapes, overlapped randomly. The effect was to interrupt perception of the stimulus pattern markedly both by interfering shapes and colors. Basically, they wanted subjects to identify objects within a noisy environment.

The subject's task was to determine on which side of the card the picture of the object appeared. 3 groups of subjects were differently presented with the audio of the superordinate, basic, subordinate categories, before being shown the pictures. They were told in advance that the target picture would appear equally often on the right and left. Subjects pressed a telegraph key with their right hand if they thought the object was on the right, and a telegraph key with their left hand if they thought it on the left.

Results / Discussion: Basic level categories aid in detecting a picture; superordinate names do not, and subordinate names aid detection no more than basic level names.

Experiment 6:

Hypothesis: Basic level names are the most abstract category names which affect "same" responses under physical identity instructions, i.e., when subjects are told to identify whether two objects are "same" or equivalent in some sense, priming with the basic level category name helps most.

Method: Subjects presented with pictures of objects belonging to the same and different categories, and told to label them as "same" or "different". They are primed with the subordinate, basic and superordinate category names.

Results / Discussion: Priming with basic level name reduces response time for “same”. Both superordinate and subordinate priming on the other hand, do not help.

Experiment 7:

Hypothesis: basic objects are first seen or recognized as members of their basic category (with additional processing required to identify them as members of their superordinate or subordinate category).

Method: Subjects were told that they would hear the name of a kind of object and immediately thereafter would see a picture. The subject was instructed to press a response key with the forefinger of his dominant hand if the picture was an object of the type named, and to press another key with his nondominant hand if the picture was not an object of that type

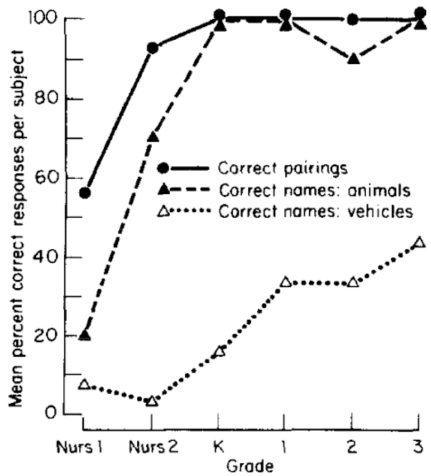
Results / Discussion: Basic level category names were faster than either superordinate or sub- ordinate names and that superordinate category names were faster than subordinate.

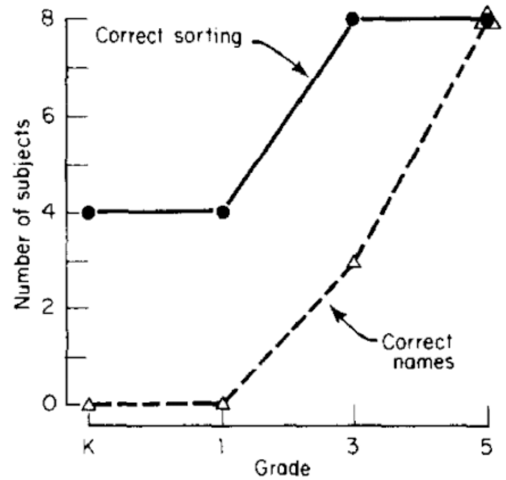
The exact nature of the additional processing required for superordinate and subordinate identifications cannot be determined from this experiment. It is speculated that superordinate identifications are derived through inference from the basic object class while subordinate identifications are derived from observation of attributes that are relevant to distinguishing between subcategories.

Experiments 8 and 9:

Hypothesis: Basic level categories are not dependent upon the development of mature adult naming or reasoning. It occurs at the earliest ages and is independent of superordinate sorting or of a child’s ability to explain the categories.

Method: Subjects at each of the following ages-3 yr, 4 yr, kindergarten, grades 1, 3, and 5 (ages 5, 6, 8, and lo), and adults were divided into two groups. One group was given an opportunity to sort sets of color pictures of common objects such as ani- mals, vehicles, clothing, and furniture into groups of basic level objects. The other group was given the same pictures, but in sets cross-cutting the basic level so that taxonomic sorting would necessarily be at the usual superordinate level. Triads of pictures were constructed and subjects had to put together the pair that was "alike"





Results / Discussion: Basic level sorts were virtually perfect; for the 3-yr-olds, the percentage correct for basic level sorts was 99, and for 4 yr and older, basic level sorts were perfect. Per- formance was considerably lower for the youngest age group, however, on sorts of the triads which could only be paired at the superordinate level: 3-yr olds, 55% correct; 4-yr olds, 96% correct.

Pictures of objects classifiable into basic level categories were classified in an adult taxonomic manner by children at all ages, including 3-yr-old children. Only the sorting of superordinate level objects showed the usual improvement with age. Evidence was presented that these results are not simply due to difference in knowledge of names for basic and superordinate level categories.

Experiment 10:

Hypothesis: object names at the basic level of abstraction should be the names by which objects are most generally designated by adult speakers of the language.

Method: Subjects were presented with pictures of objects and told to freely designate a name to it

Type of Name Given in Free Naming of Pictures

Contrast set	Type of name given			
	Superordinate	Basic level	Subordinate	Other
Superordinate	0	532	5	2
Basic level	0	533	4	2
Subordinate	1	530	5	4

Results / Discussion: There was virtually total agreement in the use of basic level names for 54 objects from nine taxonomies. It was shown that these results were not an artifact of word frequency or lack of knowledge.

Experiment 11:

Hypothesis: basic level names should be the first linguistic labels for objects acquired by the child.

Method: Corpus of 2-hour weekly recordings of Sarah's spontaneous speech during her initial period of language acquisition was collected. Two raters read the protocols and recorded all utterances of an item in any of the nine superordinate taxonomies previously studied. Utterances were classified as superordinates, basic level, or subordinates based on their level of linguistic contrast. Repetitions of an adult's or Sarah's own utterance were not included.

Results / Discussion: Both in total number of utterances of any word in a classification and in number of different words in the classification used, basic level names were es- sentially the only names used by Sarah in Stage I.

Experiment 12:

Hypothesis: if a language lacks taxonomic depth in domains of concrete objects, it can be predicted that basic level classes will be present and that it will be superordinates and/or subordinates which are lacking.

Method: The study aimed to investigate the existence of signs for different levels of abstraction in American Sign Language (ASL). The study was conducted by the fifth author, who is a fluent ASL speaker, and four informants were used, including three deaf individuals who were native ASL speakers and one hearing linguist who was fluent in ASL. The informants were interviewed extensively, and the study aimed to identify those categories for which a consistent sign or sign combination was used or which were not coded at all in ASL. A consistent sign or sign combination was considered to exist if any of the four informants thought it existed.

Results / Discussion: significantly greater percentage of signs at the basic level than at either the super- ordinate or subordinate, and that there were significantly more signs at the superordinate than at the subordinate levels.

Excavating AI

The Politics of Images in Machine Learning Training Sets

This article discusses how the computer vision field does not often think about the categorization assumed by the datasets used and how, despite what claims about the objectivity of classification algorithms people might make, such categorizations are often riddled with politics and ideology and can have serious impacts on society.

—

Training AI

- Some initial discussion on how build (supervised) machine learning algorithms requires sets of labeled images
- Labeling images is a complex, subjective process
  - "Images do not describe themselves"
  - Entire subfields of philosophy have studied the unstable relationship between images and meanings.

The Anatomy of a Training Set

- The "architecture" of a dataset can be defined by three layers
  - *Taxonomy*: "the aggregate of classes and their hierarchical nesting"
  - *Individual classes:* "singular categories that images are organized"
  - *Individually labeled images*
- Discusses the architecture of an example dataset: JAFFEE
  - *Taxonomy:* facial expressions depicting the emotions of Japanese women
  - *Classes:* happiness, sadness, surprise, disgust, fear, anger, and neutral
  - *Images:* pictures of japanese models
- There are several implicit assertions at the various layers of JAFFEE's architecture
  - Emotions can be fully captures visually, and in particular, from pics of faces
  - There is such a thing as a "neutral" face, or that these 6 emotions are complete categorization of human emotions
  - That the images represent women *feeling* those emotions rather than models acting them

- The article then moves onto discussing ImageNet (probably the most significant dataset in CV history)
  - A collection of 14M labeled images with 20K categories, released in 2009
  - Collected by scraping images and asking crowd works to label them
  - Was critical to the success of modern CV systems

- Starts by discussing the *taxonomy* of ImageNet
    - Based on semantic structure of WordNet
        - Lexical database of words and their relationship to each other
        - Attempts to categorize the entire English language
    - Attempts to categorize images of *nouns*
        - Nouns are things that pictures can represent
        - Hierarchical categorization
- Authors argue that all taxonomies/categorizations are political
    - female/male body under natural object, hermaphrodite under person
- Some discussion of the problem of categorization
    - *"To impose order onto an undifferentiated mass, to ascribe phenomena to a category—that is, to name a thing—is in turn a means of reifying the existence of that category."*
    - Some nouns are more *"nouny"* than others
        - Apple > Light > Health
    - ImageNet makes no distinction between the "nouniness" of the categories
- The authors then discuss a particular problematic top-level category: **Person**
    - Classify people into a vast range of types
        - race, nationality, profession, behavior, character, morality, etc...
    - Certain categories are problematic
        - bad person, call girl, drug addict, loser, pervert, ...
    - The collection process was also problematic
        - people's selfies and vacation photos without their knowledge

- An interesting side-note: *"ImageNet Roulette: An Experiment in Classification"*
    - Trained classifier solely on **person** category of ImageNet
    - Upload a picture, get categorized!
    - Link doesn't work anymore 🙁

- Also discusses the problem of mislabelling, particularly with humans
    - oversimplify and compress images into banalities
    - labels are nonsensical and derogatory.
    - ImageNet assumes a person's character can be determined by observing features of their bodies and faces.
- Some discussion about two other datasets
    - *UTK: Making Race and Gender from Your Face*
        - Image of people + race, gender and age
    - *IBM'S Diversity in Faces*
        - Instead of traditional labels, pictures are annotated for facial symmetry and skull shapes
        - Weirdly close to phrenology

Epistemics of Training Sets

- CV systems assume (1) fixed and universal concepts and (2) correspondences between images and concepts.
    - The existence of a "prototype" for each class, and that the prototype is visual
- The training sets used in AI are built on unstable epistemological and metaphysical assumptions, and are a form of politics.
    - There is no neutral, natural, or apolitical vantage point that training data can be built upon.
- Discussions around bias in AI systems miss the mark because there is no easy technical fix by simply shifting demographics or deleting offensive terms.

Missing Persons

- The article also discusses how some vision datasets with problem categorizations of people have "disappeared"
    - ImageNet's Persons category
    - MS-CELEB
    - Duke University's surveillance-camera footage dataset
    - Aerial Violent Individual (AVI) (exaggerated imitations of violent acts)
- However removing these datasets doesn't solve the problems they introduced
    - Have been downloaded countless times and integrated into many production AI systems
    - Removing them entirely can hinder researchers' understanding of how AI systems work and replicate biases

—

The article concludes arguing that the use of training sets in AI is not neutral, as they carry with them the biases, values, and ideologies of their creators and the society that produced them. The development of AI training sets is not just a technical issue, but also a political one, as it can shape the world in ways that reflect the interests of those with the power to create and use them.

—

Follow Questions:

- It seems that a lot of this discussion is focused on the paradigm of *supervised learning*. Does the current paradigm of *unsupervised* pre-training help alleviate problems with categorization?
- While datasets that categorize people will probably always be riddled with politics, is this really an issue with datasets that are used purely for object recognition? What other examples of problematic categorization exist?