**Epistemic Annotation Guidelines 11.16.18**

**Overview**

The goal of this work is to identify and classify epistemic statements in the scientific literature and the general subject of such statements. An epistemic statement is a statement of unknowns, hypotheses, speculations, uncertainties, including statements of claims, hypotheses, questions, explanations, future opportunities, surprises, issues, or concerns (more clarifications in the task description below).

We identify these statements using epistemic cues – how these unknowns, hypotheses, speculations, and uncertainties are communicated in text. Generally, statements of fact or methodology are not epistemic statements because the statements tend to be certain or explain the experimental design. This is not to say that there are never epistemic statements in the methods sections (see the second abstract example). For each epistemic statement, there is also the subject of it. Here that includes statements about: function, global organization, location/context, production, structures/entities, or timing.

The task at hand has 3 parts: Given a sentence (1) identify and classify epistemic cues that signify a specific type of epistemics (2) identify and classify the subject of the specific types of epistemics (3) link the epistemic cues with its subject. The end goal of this work is to determine goals for knowledge (next steps or questions) based on the epistemic statements to provide an overview of the questions in a field and to drive literature-based discovery (finding relevant articles to the goals for knowledge).

1. **TASK AT HAND**

You will be provided with 54 numbered sentences in one document in knowtator2. I will train you in using knowtator2. As you go along please take notes in the last sheet of the excel spreadsheet provided for the epistemic type, epistemic subject, how hard the sentence was to classify, and any additional notes.

For each sentence please complete all 3 tasks at hand:

1. Identify and classify spans of text that are epistemic cues that signify a specific type of epistemics. The different types along with the definition, a knowledge goal, and linguistic cues that signify the specific type are provided. Each linguistic cue includes at least one positive and negative example with the specific linguistic cue highlighted in yellow and the scope in parentheses <…> for positive examples and highlighted in red for negative examples. Negative examples will always be last. Linguistic cues can be disjoint, meaning separated by other words – we write disjointness as ‘…’. Explanations for the annotation are in parentheses […] in orange. All examples are from the PubMed Central Open Access Subset.

Feel free to use this to ensure it is an epistemic statement with more information needed (next steps). Further, within knowtator2 there are synonyms and you can use the search function to help navigate within knowtator2. Also, if you think I am missing a type of epistemics please add it to the last sheet in the excel document under “Epistemic Type Notes”.

* 1. **Claims/Questions/Evidence**
     1. Full unknown = a statement that indicates something is not known (a lack of information), or information is presented for the first time (new or novel); not a result statement about the absence of something.
        1. **Knowledge goal:** explore the unknown further to gain any insights
        2. **Linguistic cues:** with positive and negative examples
           1. **LACK OF…DATA**

“despite the rapid increase in the extent of research into isoflavones in recent years, there remains a <lack of data on which to base recommendations regarding intakes either for patients with breast cancer or healthy women wishing to reduce their risk of breast cancer>.” [lack of information to understand recommendations]

“the following conditions were exclusion criteria for recruitment: the indicators and ctc were not related; the evaluation indicators derived from ctc in blood circulation were not used for hcc diagnosis; lack of complete data to describe or calculate the sensitivity and specificity; and reviews, letters, technical reports, case reports, and comments.” [a lack of data is an exclusion criterion only]

* + - * 1. **LITTLE IS…KNOWN/ VERY LITTLE IS…KNOWN**

“<little is known about the feeding behavior of hematophagous insects that require plant sugar to complete their life cycles>.” [almost nothing is known about feeding behavior]

“<little information is known about the biological activities of these metabolites>.” [almost nothing is known about the biological activities of these metabolites]

NO NEGATIVES [almost always a positive example indicating information is lacking]

* + - * 1. **NO… AVAILABLE/ NOT…AVAILABLE**

“to date there are <no therapies available that modify disease progression in plan, pkan or other forms of nbia>.” [there is a lack of therapies to modify disease progression]

“the largest positive sensitivities are found for a9and a10, whose values were set equal (sees2 table) as <no data on the phosphorylation of pdk1 and mtorc2 were available>.” [there is a lack of data on phosphorylation]

“in lieu of antibodies capable of detecting native endogenous perforin-2, <as these reagents are not currently available>, we transfected c-terminal tagged perforin-2–rfp or–gfp into cells in which endogenous perforin-2 was knocked down.” [there is a lack of reagents]

“in case of doubts or new symptoms, patients are instructed to access this consultation, instead of emergency room, where misinterpretation of the symptoms is more frequent, particularly when neurologist is not available.” [there is a lack of neurologist for a patient, not a lack of information]

“for some patients, complete clinical data were not available.” [specific patients are missing data which is not a research question]

* + - * 1. **NOT…CLEAR**

“<although the pathogenesis of bisphosphonate-induced esophageal mucosal damage has not been clearly demonstrated>, direct chemical esophageal damage with prolonged local mucosal exposure to a drug with gastric acid might be the most plausible biological mechanism as suggested by the literature.” [the pathogenesis is not known]

“if the statistical methods used in a biomarker study are not clearly specified, it will be difficult or impossible for the reader to interpret the results or reproduce and validate the findings.” [the statistical methods need to be specified but they are not unknown]

* + - * 1. **NO…EXIST/NO…EXIST**

“<while such technologies for remote sensing do not yet exist>, it may be possible to use such methods in the future to detect diseases without clear visible symptoms.” [technologies are not known since they don’t exist – a new methodology is necessary]

“faf of rpe tears showed a hypofluorescent area in the rpe denuded zone, indicating non-existing rpe.” [results are explained that rpe is missing but not that it is unknown]

“no existing method other than mrna delivery could achieve the direct, vector-independent expressions of transcription factors in an exogenous manner.” [no exist signifies hyperbole that there is only one method currently to use – not that a method doesn’t exist]

* + - * 1. **NO…IS KNOWN**

“<nothing further is known about distances travelled to form a new nest>.” [a lack of information about distances traveled to form a new nest]

“to our knowledge, <not much is known about the injury pattern of certain injury mechanisms in different groups of young road users (3–22 years) compared to a large group of adult road users>.” [a lack of information about injury patterns in specific populations]

“virtually <nothing is known, however, on the possible links between disturbance and epigenetic features of wild plant populations>.” [a lack of information on the links between disturbance and epigenetic features]

“variables with finite variance approximately follows a normal distribution), no matter whether the distribution of sound correspondences in v is known or not…” [the distribution is known with or without any information – a statement of knowledge not being necessary]

* + - * 1. **NO…IDENTIFIED**

“interestingly, we have <not identified a specific relation between serum dkk-1 concentrations and patterns of bone pathology in psa, in the form of either new bone formation or bone loss>.” [the specific relation between serum dkk-1 concentrations and patternsof bone pathology is unknown]

“t tumour, n inguinal lymph node, 0 no such cells identified, (+) present, (−) absent” [no…identified defines 0 here so nothing is actually unknown]

“this amino-acid substitution was not identified in rr mice, in concordance with the absence of cho qtl in b6 × rr f2mice.” [a statement of something absent not unknown”

* + - * 1. **COULD NOT DETERMINE**

“we <could not determine the superiority of la peak systolic strain over la fractional shortening>. [the superiority is unknown]

“reasonable care and skill could not have remedied the accident, and the lawyers could not determine what they were sueing for until after a dissection of their client.” [the lawyers don’t know but there is no general research question that is unknown]

* + 1. Explicit questions = an explicit statement of inquiry (with a question mark or question word such as how, where, what, why)
       1. **Knowledge goal:** find answers to the question; discover methodologies that will help answer the question
    2. Alternative options = a statement of multiple (at least 2) choices or actions with no stated disagreements, including statements with an implied second option, such as “whether”
       1. **Knowledge goal:** determine the correct option or a better option
       2. **Linguistic Cues:** with positive and negative examples
          1. **AND**
    3. Incomplete evidence = a statement of one proposed explanation for a phenomenon on the basis of limited evidence as a starting point for further investigation OR a statement that information is needed to support an assertion or claim, including both positive and negative statements. Either a statement that some evidence already exists, explaining how current findings support previous work, adding confidence to a claim OR a statement that information is limited, more research is needed or is ongoing including limitations – biases or short comings related to the study design and execution.
       1. **Knowledge goal:** gather more evidence to support the claim OR conduct more research to determine the validity of the claim; Complete the partial picture; Consider the short comings and biases for the next experiment and how it can be addressed
  1. **Links/Understanding**
     1. Possible understanding = a statement of a feasible explanation, relationship, or phenomenon; no claim to the correct explanation, relationship or phenomenon
        1. **Knowledge goal:** determine the most feasible or correct explanation, relationship, or phenomenon either mentioned or not
     2. Probable understanding = a statement staking a claim to the most likely explanation, relationship, or phenomenon; there is a good chance this understanding is correct
        1. **Knowledge goal:** determine if the most likely option is correct of if it’s another option
     3. Superficial relationship = a statement about a connection, link, or association between at least 2 variables; connectedness between entities and/or interactions representing their relatedness or influence
        1. **Knowledge goal:** confirm the connection, link, or association between variables; determine the full underlying relationship between variables
     4. Etiology = a statement about an incomplete underlying cause or mechanism
        1. **Knowledge goal:** complete and discover the missing underlying cause or explanation
  2. **Future opportunities**
     1. Future work = a statement of extensions, including next steps, directions, opportunities, approaches, or considerations of the described work that may be implemented at some future time point
        1. **Knowledge goal:** determine the next course of action based on this future work proposal
     2. Non-urgent recommendation = a statement of suggestion or proposal as to the best course of action, especially one put forward by an authoritative body OR advice telling someone what the best thing to do is
        1. **Knowledge goal:** determine the validity of the suggestion or proposal; publicize the suggestion or proposal to the correct group of people
     3. Future prediction = a statement of extrapolation of given data into the future and/or from past observations; no reference to next steps
        1. **Knowledge goal:** run the simulation or experiment to determine if the prediction is correct; publicize the outcomes of the study to the correct people
     4. Urgent call to action = a statement calling for immediate attention including an action needed to be taken immediately or information that needs to be disseminated immediately OR critical: being in or verging on a state of crisis or emergency OR urgently needed OR absolutely necessary
        1. **Knowledge goal:** take the urgent action ASAP or distribute the knowledge ASAP
  3. **Surprises/Interesting**
     1. Curious finding = a statement of a surprising result, conclusion, or situation; the researchers were not expecting the result, conclusion or situation but are intrigued by it
        1. **Knowledge goal:** explore the surprising result, conclusion, or situation more
     2. Unexpected observation = a statement of a judgement or inference from an observable phenomenon that the researchers did not expect; an element of surprise
        1. **Knowledge goal:** repeat the experiment or simulation to make sure the observations are repeatable; think about next steps with this result
  4. **Issues/concerns**
     1. Difficult task = a statement of something not easily done, accomplished, comprehended, solved, or complicated with a multitude of underlying pieces or parts; heterogeneity; not medical complications
        1. **Knowledge goal:** create methods to study the complicated system and to better understand any piece of the complicated system; potentially overall field development needed
     2. Problems or complications = a statement of issues, problems, mistakes, or medical complications that are cause for anxiety and/or worry
        1. **Knowledge goal:** determine the gravity of the concern and determine if it needs to be dealt with before the next experiment or study
     3. Controversy = a statement of disagreement amongst researchers OR a lack of consensus OR at least two possible answers are presented as results from different researchers - usually in reference to previous results and stated when results disagree with each other OR contradictions
        1. **Knowledge goal:** determine the truth to break any disagreements
  5. **Questions answered by this work** = a statement of a goal or objective of a study that is attempted or completed during the study
     1. **Knowledge goal:** find the answer(s) in the article; determine if the question(s) is (are) fully answered in the article

1. Identify and classify the span of text that is the subject of the type of epistemics: what is the epistemic statement about in general (full definitions of subjects can be found in attached excel spreadsheet - Ignorance\_Taxonomy\_Information\_10.8.18 – Epistemic\_Subjects Sheet). Along with the definition, a knowledge goal is provided. Feel free to use this to ensure the annotating captures that more information is needed (next steps). Each epistemic type identified in task 1 should have a corresponding subject. If you do not see one that fits, please think about a new one that would capture the subject and write a note in the last excel spreadsheet under “Epistemic Subject Notes”.
   1. **Function**
      1. Abilities = the skills and potential to act
         1. **Knowledge goal:** What can things do/not do in the relevant circumstances?
      2. Activities = the realization of abilities
         1. **Knowledge goal:** What actions occur at different time points?
      3. Roles = the reason for the activity or entity with respect to the system
         1. **Knowledge goal:** What is the functional role of the entity or activity in the context of the system? What does it contribute to the behavior of the system as a whole?
   2. **Global Organization**
      1. Communication = contact between at least two things
         1. **Knowledge goal:** Do these things communicate with each other? How do they communicate?
      2. Comparison = determine the similarities and differences between at least two things
         1. **Knowledge goal:** How are these things similar and different?
      3. Disruption = disturbances, problems, or alterations that interrupt an event, activity, or process
         1. **Knowledge goal:** What is disrupted? How is it disrupted?
      4. Independent = at least two things are disjoint
         1. **Knowledge goal:** Are these things independent? Why are they independent?
      5. Pathway = a path between at least two things
         1. **Knowledge goal:** What is the path between at least two things?
      6. Regulation = constraints on entities or processes from outside the process or entity itself, including preventative measures
         1. **Knowledge goal:**  How is the process regulated?
      7. Reproducibility = multiple sources/studies have come to the same conclusion or results
         1. **Knowledge goal:** Are the same results or conclusions reached when the study or a similar one is done?
      8. Risk = either a chance with a possibility of being wrong or right with no guarantee either way or hazardous/potentially harmful/no guarantee of benefits
         1. **Knowledge goal:** Is something a real risk? How big of risk is it?
      9. Safety = efficacy concerns or focus
         1. **Knowledge goal:** Is the experiment safe?
      10. Model = the overall large picture organization of a system
          1. **Knowledge goal:** Does the model work with all situations mentioned? Are there any exceptions or edge cases?
      11. Methodology = instrumentation or methodologic information
          1. Is the instrumentation appropriate? Does it fulfill all its needs?
   3. **Location/Context**
      1. Body location = located in a specific part of the organism (like a body part or organ) instead of referring to the whole organism
         1. **Knowledge goal:** Where in the organism?
      2. Environment = the outside surroundings of the system
         1. **Knowledge goal:** What are the surroundings? What role do the surroundings play?
      3. Model/animal system = the whole organism
         1. **Knowledge goal:** What organism?
      4. Sex = societal sex assignment (male or female typically)
         1. **Knowledge goal:** What is the gender/sex? Does the gender/sex make a difference in the system?
   4. **Production**
      1. Establishes = information about establishment including formation and signaling
         1. **Knowledge goal:** What establishes it? How is it established? What forms it? What signals it?
      2. Mobilization = movement and/or usage of stored material waiting for use
         1. **Knowledge goal:** Why are the stores mobilized now?
      3. Sustains = information about its maintenance including reduction, loss, and retention
         1. **Knowledge goal:** What sustains it? Is it sustained?
   5. **Structures/Entities**
      1. Characteristics = an other category for structures/entities
         1. **Knowledge goal:** What are structural and entity characteristics?
      2. Magnitude/amount/size = the size, extent, or amount
         1. **Knowledge goal:** How big? How much?
      3. Molecular nature = genetic and chemical underpinnings of the structure or entity
         1. **Knowledge goal:** What is the molecular nature?
      4. Morphology = the form of the entity or structure
         1. **Knowledge goal:** What is the morphology?
      5. Structure, consistency, uniformity = the overall stability of the structure or entity over time and location
         1. **Knowledge goal:** Is the structure consistent or uniform?
   6. **Timing**
      1. Age = length of time a thing has existed
         1. **Knowledge goal:** What is the age?
      2. Duration = the length of time for a process
         1. **Knowledge goal:** How long is the process?
      3. End time = the termination time of a process
         1. **Knowledge goal:** When does the process end?
      4. Start time = the beginning time of a process
         1. **Knowledge goal:** When does the process start?
      5. Time points = specific time points during the process
         1. **Knowledge goal:** When does a specific event occur?
2. Link the type of epistemics (task 1) to its subject (task 2). Using the graph space functionality in knowtator2, connect the epistemic types to their subjects using the relation “is about”. I will provide detailed training of how to use knowtator2 for this task on the day of annotation. Please put all annotations from one sentence in the same graph space even if parts are disconnected. Name the graph space GS# where # is the sentence number.

**NOTES:**

* 1. There can be no, one, or more than one type of epistemics per sentence. There can also be no, one, or more than one subject per sentence. Types of epistemics and subjects can form a chain with the relation is\_about. Please also put all annotations from one sentence in the same graph space even if parts are disconnected.
  2. Each type of epistemics should either map to or be mapped to either a subject or another type of epistemics. Graphing the “is\_about” relationship helps clarify this criterion.
  3. Feel free to use the search function built in to knowtator2: make sure to read the options and be careful because terms can appear in multiple categories.

1. **2 ABSTRACT IDENTIFICATION EXAMPLES**

**Key: Yellow is epistemic type (TYPE), green is subject (SUBJECT)**

* 1. **PMC48141 abstract**

**Background**

Little is known (FULL UNKNOWN) about genetic factors (MOLECULAR NATURE) affecting intraocular pressure (IOP) in mice and other mammals (MODEL/ANIMAL SYSTEM). The purpose of this study (QUESTIONS ANSWERED BY THIS WORK) was to determine the IOPs of genetically distinct mouse strains (MODEL/ANIMAL SYSTEM), assess the effects of factors such as age (AGE), sex (SEX) and time of day (TIME POINT) on IOP in specific strain backgrounds, and to assess the effects of specific candidate gene mutations (MOLECULAR NATURE/DISRUPTION) on IOP.

**Results**

Based on over 30 studied mouse strains, average IOP ranges from approximately 10 to 20 mmHg. Gender (SEX) does not typically (CONTROVERSY) affect IOP and aging (AGE) results in an IOP decrease in some strains (MODEL/ANIMAL SYSTEM). Most (PROBABLE EXPLANATION) tested strains (MODEL/ANIMAL SYSTEM) exhibit a diurnal rhythm with IOP being the highest during the dark period of the day (TIME POINT). Homozygosity for a null allele of the carbonic anhydrase II gene (*Car2**n*) does not alter IOP while homozygosity for a mutation in the leptin receptor gene (*Lepr**db*) that causes obesity and diabetes results in increased IOP. Albino C57BL/6J mice homozygous for a tyrosinase mutation (*Tyrc*-2*J*) have higher IOPs than their pigmented counterparts.

**Conclusions**

Genetically distinct mouse strains housed in the same environment have a broad range of IOPs. These IOP differences are likely (PROBABLE EXPLANATION) due to interstrain (MODEL/ANIMAL SYSTEM) genetic differences (MOLECULAR NATURE) that create a powerful resource for studying the regulation (REGULATION) of IOP. Age, time of day, obesity and diabetes have effects on mouse IOP similar to those in humans and other species. Mutations in two of the assessed candidate genes (*Lepr* and *Tyr*) result in increased IOP. These studies demonstrate that mice are a practical and powerful experimental system to study the genetics of IOP regulation and disease processes that raise IOP to harmful levels.

* 1. **PMC5975408**

## **Background**

Prevention of mother-to-child HIV transmission (PMTCT) programs usually (INCOMPLETE EVIDENCE) test pregnant women (ANIMAL/MODEL SYSTEM) for HIV without involving (INDEPENDENT) their partners. Non-disclosure of maternal HIV status to male partners (COMMUNICATION) may (POSSIBLE UNDERSTANDING) deter (DISRUPTION) utilization of PMTCT interventions since partners play a pivotal role (ROLE) in decision-making within the home including access to and utilization of health services.

## **Methods**

Mothers attending routine 6-week and 9-month infant immunizations were enrolled at 141 maternal and child health (MCH) clinics across Kenya from June–December 2013. The current analysis was restricted to mothers with known HIV status who had a current partner. Multivariate logistic regression models adjusted for marital status, relationship length and partner attendance at antenatal care (ANC) were used to determine correlates of HIV non-disclosure among HIV-uninfected and HIV-infected mothers, separately, and to evaluate the relationship (SUPERFICIAL RELATIONSHIP) of non-disclosure with uptake of PMTCT interventions (COMMUNICATION). All analyses accounted for facility-level clustering,

## **Results**

Overall, 2522 mothers (86% of total study population) met inclusion criteria, 420 (17%) were HIV-infected. Non-disclosure of HIV results to partners was higher among HIV-infected than HIV-uninfected women (13% versus 3% respectively, p < 0.001). HIV-uninfected mothers (MODEL/ANIMAL SYSTEM) were more likely (PROBABLE UNDERSTANDING) to not disclose (COMMUNICATION) their HIV status to male partners if they were unmarried (adjusted odds ratio [aOR] = 3.79, 95% CI: 1.56–9.19, p = 0.004), had low (≤KSH 5000) income (aOR = 1.85, 95% CI: 1.00–3.14, p = 0.050), experienced intimate partner violence (aOR = 3.65, 95% CI: 1.84–7.21, p < 0.001) and if their partner did not attend ANC (aOR = 4.12, 95% CI: 1.89–8.95, p < 0.001). Among HIV-infected women (MODEL/ANIMAL SYSTEM), non-disclosure (COMMUNICATION) to male partners was less likely (PROBABLE UNDERSTANDING) if women had salaried employment (aOR = 0.42, 95%CI: 0.18–0.96, p = 0.039) and each increasing year of relationship length (DURATION) was associated (SUPERFICIAL RELATIONSHIP) with decreased likelihood (PROBABLE UNDERSTANDING) of non-disclosure (COMMUNICATION) (aOR = 0.90, 95% CI: 0.82–0.98, p = 0.015 for each year increase). HIV-infected women who did not disclose (COMMUNICATION) their HIV status to partners were less likely (PROBABLE UNDERSTANDING) to uptake CD4 testing (aOR = 0.32, 95% CI: 0.15–0.69, p = 0.004), to use antiretrovirals (ARVs) during labor (OR = 0.38, 95% CI 0.15–0.97, p = 0.042), or give their infants ARVs (OR = 0.08, 95% CI 0.02–0.31, p < 0.001) (DISRUPTION).

## **Conclusion**

HIV-infected women (MODEL/ANIMAL SYSTEM) were less likely (PROBABLE UNDERSTANDING) to disclose (COMMUNICATION) their status to partners than HIV-uninfected women (MODEL/ANIMAL SYSTEM). Non-disclosure (COMMUNICATION) was associated (SUPERFICIAL RELATIONSHIP) with lower use of PMTCT services (DISRUPTION). Facilitating maternal disclosure to male partners (COMMUNICATION) may (POSSIBLE UNDERSTANDING) enhance (SUSTAINS) PMTCT uptake.

1. **CLARIFYING CONFUSING DISTINCTIONS – put this in the first one that appears and in the other one say see note above**
2. **Possible understanding vs. Probable understanding**

Possible understanding is a statement of a feasible explanation, relationship, or phenomenon with no claim to the correct explanation, relationship, or phenomenon. The statement is more proposing options and not staking a claim as to which is correct. This includes epistemic cues such as could, can be, have potential, it is possible, and might. In the sentence, “Small changes in the resistance to aqueous humor drainage may also contribute to diurnal differences in IOP” the researchers offer another possible contributor to diurnal differences in IOP.

Probable understanding is a statement staking a claim to the most likely explanation, relationship, or phenomenon where it has a good chance of being the case or of coming about. This includes epistemic cues such as it is becoming increasingly clear. Likely, unlikely, promising, plausible, mainly. In the sentence “It is becoming increasingly clear that many forms of glaucoma have a genetic component” the researchers stake a claim that probably many forms of glaucoma have a genetic component.

1. **Controversy vs. alternative options**

Controversy is a statement of disagreement amongst researchers; a lack of consensus; at least two possible answers are presented as results from different researchers; usually in reference to previous results and stated when results disagree with each other; contradictions. There are at least two possible answers already presented as results from different researchers. This is usually in reference to previous works findings and stated when findings disagree with each other. This includes epistemic cues such as do not agree, have been variably, some but not all, some but not other studies, have been implicated, and but. In the sentence “Some but not all human studies have reported a positive association between IOP and blood pressure” there is disagreement on the association between IOP and blood pressure.

Alternative options is a statement of multiple (at least 2) choices or actions with no stated disagreements, including statements with an implied second option such as "whether". These options have not yet been studied and so there cannot be disagreements yet about which one is correct. The statement only introduces the options. This includes epistemic cues such as alternatively, cannot be ruled out, though, either, and whether. In the sentence “An effect of anesthesia in these very old mice cannot be ruled out” the researcher indicates an anesthesia affect is an alternative option for explaining some phenomenon in very old mice.

1. **Curious finding vs. unexpected observation**

Curious finding is a statement of a surprising result, conclusion, or situation; the researchers were not expecting the result, conclusion, or situation but are intrigued by it. This includes epistemic cues such as curious, however, interestingly, surprisingly, and suspected. The researchers were not expecting the result or conclusion.

Unexpected observation is a statement of a judgement or inference from an observable phenomenon that the researchers did not expect; an element of surprise. This includes epistemic cues such as apparently, appear, appears to be, seems to be, and does not appear to be. The statement describes an observable phenomenon that they did not expect.

In the sentence “Interestingly, the IOP of strain CBA/CaJ does not appear to increase in the dark.” *Interestingly* is classified as curious finding because it indicates a result that the researchers did not expect. *Does not appear to* is classified as unexpected observation because they observed that the IOP not increasing in the dark but they are not certain if this observable phenomenon is repeatable.

1. **Non-urgent recommendation vs. urgent call to action**

Non-urgent recommendation is a statement of suggestion or a proposal as to the next best course of action, especially one put forward by an authoritative body; advice telling someone what the best thing to do is. This includes epistemic cues such as advocate, encourage, guideline, should be, require, must be, and it is important that. It is a non-urgent recommendation that does not necessarily need to be dealt with right away, there is no clear sense of urgency. For example, in the sentence, “Thus, to avoid effects of anesthesia on IOP, all measurements should be made within a window of up to 12 minutes after anesthetic administration.” The sentence recommends taking measurements up to 12 minutes after anesthetic administration, without urgency, but useful for further experimentation.

Urgent call to action is a statement calling for immediate attention including an action needed to be taken immediately or information that needs to be disseminated immediately; critical: being in or verging on a state of crisis or emergency; urgently needed; absolutely necessary. This includes epistemic cues such as call to act, critical, cause for concern, crucial, essential, high priority, imperative, needed, influential, and vital. The statement has a clear sense of urgency. For example, in the sentence, “As PMTCT Option B+ scales up with regimens that require life-long antiretroviral treatment (ART) adherence, the need for HIV-infected pregnant and breastfeeding women to disclose their status to their male partners becomes more critical in order to maintain adherence to ART.” Here *the need* and *more critical* indicate a sense of urgency that needs to be dealt with sooner rather than later.

1. **Future work vs. future prediction**

Future work is an explicit statement of extensions, including next steps, directions, opportunities, approaches, or considerations, of the described work that may be implemented at some future time point. This includes epistemic cues such as additional studies, more information… is needed, needs to be assessed, future investigation/strategies/opportunities/considerations, and candidate. For example, in the sentence “Mutations in two of the assessed candidate genes (*Lepr* and *Tyr*) result in increased IOP.” *Candidate* indicates that more work needs to be done to determine if these 2 genes actually increase IOP.

Future prediction is a statement of extrapolation of given data into the future and/or from past observations; no reference to next steps or future work. This includes epistemic cues such as if real, if confirmed, expect, supposed to, will, ultimately, would, and implication. For example, the sentence “If real, this sporadic sex difference was not dependent on age, sometimes occurring in a group of B6 mice at a particular age and sometimes not occurring in a separate group of the same age.” This sentence indicates that if the phenomenon was real, then multiple things follow (a prediction). Notice no next steps are stated.