

ОБЩАЯ ПСИХОЛОГИЯ, ПСИХОЛОГИЯ ЛИЧНОСТИ, ИСТОРИЯ ПСИХОЛОГИИ

ИНСАЙТ НА ВИДЕО: ЭКСПЕРТНЫЕ И НАИВНЫЕ ОЦЕНКИ ПОВЕДЕНЧЕСКИХ ПРИЗНАКОВ ИНСАЙТНОГО РЕШЕНИЯ

© Кулиев Н.С.

аспирант, Институт психологии РАН, Москва, Россия
kuliev.nikolaj@bk.ru, ORCID: 0009-0004-1003-464X

© Валуева Е.А.

кандидат психологических наук, Институт психологии РАН, Москва, Россия
valuevaea@ipran.ru, ORCID: 0000-0003-3637-287X

© Куприянов Р.В.

кандидат психологических наук, Казанский федеральный университет, Казань, Россия
rvkupriyanov@kpfu.ru, ORCID: 0000-0001-9794-9607

Исследование выполнено при финансовой поддержке Российского научного фонда (РНФ),
проект № 23–18–01059-П

Для цитирования:

Кулиев Н.С., Валуева Е.А., Куприянов Р.В.
Инсайт на видео: экспертные и наивные
оценки поведенческих признаков инсай-
тного решения // Ученые записки Ин-
ститута психологии Российской акаде-
мии наук. 2026. Т. 6. № 2(19). С. 35-48.
DOI: 10.38098/proceedings_2026_06_02_04
Kuliev N.S., Valueva E.A., Kupriyanov R.V.
Capturing insight on video: expert and naive
evaluations of behavioral markers in insight-
ful problem solving. Proceedings of the Insti-
tute of Psychology of the Russian Academy
of Sciences. 2026, Vol. 6, No2(19), Pp. 35-48.
DOI: 10.38098/proceedings_2026_06_02_04

В статье проверяется возможность объективной видеорегистрации инсайтных решений по коротким видеофрагментам поведения решателей. Эксперты оценивали 90 видео по ряду поведенческих шкал и классифицировали решения как инсайтные или неинсайтные. Далее та же логика была масштабирована на 976 видеоклипов с привлечением наивных оценщиков и калибровкой на эталонных примерах. Показано, что межэкспертная согласованность при идентификации инсайта является умеренной, но существенно возрастает в более согласованной подгруппе экспертов. Наиболее надежно распознаются мимика, удивление и позитивные эмоции. Экспертные оценки поведенческих проявлений образуют двухфакторную структуру, в которой решение о наличии инсайта в основном определяется фактором позитивной экспрессии. Агрегированные оценки наивных испытуемых в среднем достаточно хорошо воспроизводят экспертные решения, однако в большей степени опираются на общую заметность реакции, чем на дифференцированное распознавание специфических маркеров инсайта.

При этом связь как экспертных, так и агрегированных поведенческих оценок с субъективными самоотчетами решателей остается слабой. Полученные результаты показывают перспективность видеоанализа и калиброванной коллективной разметки для исследования инсайта, но одновременно подчеркивают необходимость осторожной интерпретации внешних поведенческих критериев.

Ключевые слова: инсайт, ага-переживание, самоотчет, видеоанализ, поведенческие маркеры, экспертная оценка, наивные оценщики, коллективная оценка, краудсорсинг

Введение

Инсайт традиционно определяется как специфический тип решения задачи, характеризующийся внезапностью нахождения ответа, субъективным переживанием «Ага!», ощущением очевидности и высокой уверенности в правильности решения (Bowden, Jung-Beeman, 2003; Kounios, Beeman, 2014). Современные подходы отмечают многокомпонентность

этого феномена, помимо внезапности, в его структуру включаются аффективный компонент (положительные эмоции, облегчение), метакогнитивный компонент (чувство «теплоты», переживание «Ага!», чувство очевидности и уверенности) (Морошкина и др., 2020; Чистопольская и др., 2021), а также когнитивный — репрезентативное переструктурирование задачи. При этом подчеркивается, что когнитивное

изменение и субъективное Ага-переживание тесно связаны, но не полностью тождественны друг другу (Danek, Williams, Wiley, 2020). Инсайтные решения обычно точнее аналитических (Salvi et al., 2016), однако возможны как верные решения без яркого аффекта, так и «ложные инсайты» — переживание озарения при неверном ответе (Danek, Wiley, 2017). В связи с этим распознавание инсайта остается методологически сложной задачей.

Трудности детекции инсайта обусловлены как вариативностью феноменологических проявлений, так и влиянием контекстуальных факторов. Экспериментальные условия (инструкция, способ регистрации ответа, формулировки шкал) существенно модифицируют частоту и интенсивность сообщаемых инсайтных переживаний (Laukkonen, Tangen, 2018; Лазарева, Савинова, Чистопольская, 2023). Помимо этого на точность самоотчетов влияют текущее состояние, мотивация, метакогнитивные способности, склонность к рефлексии, а также индивидуальные представления об инсайте и ожидания относительно того, каким должен быть инсайт (Шумилов, Чистопольская, 2023; Chesebrough, Oh, Kounios, 2024; Морошкина и др., 2020; Чистопольская и др., 2021). Поэтому исследователи стремятся формировать у участников единое понимание инсайта, в частности с помощью обучающих видео (Shumilov et al., 2023).

Субъективные и объективные методы регистрации инсайтных решений

Исследователи выделяют две основные группы методов регистрации инсайтных решений — субъективные и объективные (Морошкина и др., 2020; Чистопольская и др., 2021; Гаршина, Пискарев, 2023). Субъективная регистрация инсайта опирается на самоотчеты — бинарную классификацию решений как инсайтных или аналитических, а также рейтинговые шкалы для оценки внезапности, уверенности и эмоциональной окраски переживания «Ага!» (Bowden, Jung-Beeman, 2003; Ishikawa, Toshima, Mogi, 2019). Для более детального отражения феномена инсайта исследователи применяют дополнительные показатели субъективного опыта — радость, удивление, облегчение и азарт (Danek et al., 2014; Danek, 2018). В отечественных работах предложены новые шкалы самооценки **инсайтности**, которые учитывают ограничения классического опросника Данек и обеспечивают более точную и всестороннюю фиксацию переживания (Shumilov et al., 2023). Используются также визуальные самоотчеты: участникам предлагается выбрать одно из четырех графических изображений, иллюстрирующих разные дина-

мики процесса решения (Spiridonov et al., 2021). Самоотчеты дают прямой доступ к феноменологическому аспекту решения, который часто недоступен внешнему наблюдению, подробно описывают ход мыслей и легко применимы. Однако они страдают от вариативности интерпретации шкал и ненадежности воспоминаний, поэтому не могут служить единственным критерием регистрации инсайтных решений (Чистопольская и др., 2021).

Объективные методы направлены на регистрацию поведенческих, нейрофизиологических и соматических коррелятов инсайта. Нейрокогнитивные исследования связывают ага-переживание с особенностями правополушарной семантической активации и специфическими паттернами мозговой активности, возникающими до и в момент решения (Bowden, Jung-Beeman, 2003; Kounios, Beeman, 2014). Физиологические работы показывают, что инсайт сопровождается измеримыми телесными и аффективными изменениями: ЭЭГ и другие показатели позволяют фиксировать эмоциональные корреляты инсайта, сила сжатия динамометра в момент ага-переживания связана с точностью решения, а повышение кожно-гальванической реакции отражает сочетание положительного аффекта и субъективного чувства внезапной ясности (Cernea, Kerren, Ebert, 2011; Laukkonen et al., 2021; Shen et al., 2018). Окуломоторные исследования также показывают, что перед инсайтным решением меняется распределение внимания: участники чаще отводят взгляд от внешнего стимула, меняют паттерны моргания и демонстрируют расширение зрачка, что отличает инсайт от аналитического решения (Knoblich, Ohlsson, Raney, 2001; Salvi et al., 2015; Salvi et al., 2020). В качестве объективного индикатора приближения инсайта также может использоваться вторичное задание-зонд, позволяющее оценивать текущую загрузку рабочей памяти в процессе решения. Более быстрое время реакции на зонд во время инсайтного решения интерпретируется как признак меньшей вовлеченности сознательных систем обработки информации по сравнению с аналитическим решением (Korovkin et al., 2018).

Несмотря на то, что объективные методы регистрации инсайта обеспечивают высокую степень независимости от субъективных интерпретаций, точность и сопоставимость данных, они требуют специализированного оборудования и стандартизации процедур, являются затратными, нарушают естественность процесса решения и нередко дают косвенные показатели, нуждающиеся в интерпретации (Чистопольская и др., 2021; Adler, 2024).

Отдельное направление связано с анализом видеозаписей поведения участников. Этот под-

ход позволяет фиксировать внешние проявления инсайтного решения без прямого вмешательства в процесс. По данным отечественных работ, инсайт может сопровождаться различными изменениями мимики и жестов; в числе возможных маркеров обсуждаются поднятие бровей, улыбка и некоторые паттерны движений, хотя их диагностическая ценность зависит от контекста и процедуры разметки (Филяева, Коровкин, 2015; Владимиров, Бушманова, Макаров, 2021; Владимиров, Бушманова, Курицын, 2023). В этом смысле видеопроанализ занимает промежуточное положение между самоотчетом и инструментальными методами: он менее инвазивен, чем физиологическая регистрация, но требует четких критериев интерпретации и надежной процедуры кодирования (Adler, 2024).

Коллективная оценка творческих решений

В статье «Vox Populi» (1907), опубликованной в журнале *Nature*, Фрэнсис Гальтон описал конкурс по оценке массы быка, проведенный на сельскохозяйственной выставке. В исследовании приняли участие 787 человек, включая как специалистов сельскохозяйственного профиля, так и неспециалистов. Результаты показали, что несмотря на значительный разброс индивидуальных оценок их совокупность образовала распределение, близкое к нормальному, при этом центральное значение оказалось поразительно близким к истинной массе животного. Таким образом, невзирая на простоту агрегирования индивидуальных оценок, этот результат заложил основу концепции «мудрости толпы» (Galton, 1907). Современные обзорные работы показывают, что коллективный интеллект зависит не только от числа участников, но и от того, насколько группа сохраняет разнообразие индивидуальных суждений и как именно агрегирует их в общее решение (Kameda, Toyokawa, Tindale, 2022).

В рамках метода консенсусной оценки ключевым остается вопрос о том, кто именно может выступать в роли оценщика креативного продукта. В классическом варианте этого подхода наиболее валидными считаются согласованные суждения экспертов, обладающих признанной компетентностью в соответствующем домене. Вместе с тем последующие исследования показывают, что требуемый уровень компетентности зависит от типа продукта, особенностей домена и степени структурированности критериев оценки (Kaufman et al., 2008; Kaufman, Baer, Cole, 2009). Сравнительные данные показывают, что неспециалисты, как правило, не являются полноценной заменой экспертам, поскольку их оценки обычно менее согласованы и не всегда воспроизводят

экспертное ранжирование. Однако промежуточные группы оценщиков, обладающие предметной подготовкой, но не имеющие статуса признанных экспертов, в ряде случаев демонстрируют приемлемую согласованность и могут рассматриваться как квази-эксперты (Kaufman et al., 2013; Tsai, 2016). Дополнительные основания для расширения круга оценщиков дают исследования, показывающие, что при использовании более явных критериев и предварительного обучения наивные оценщики способны демонстрировать высокую межсубъектную согласованность, хотя вопрос их полной эквивалентности экспертам остается открытым (Crompton, Kaufman, 2012; Storme et al., 2014). В этом смысле привлечение наивных испытуемых целесообразно рассматривать как экономичный инструмент предварительной сортировки, агрегированной оценки или калиброванной коллективной разметки, применимость которого должна проверяться отдельно для каждого типа задачи.

Большинство существующих работ фокусируется на дивергентных задачах или на задачах, связанных с созданием творческих продуктов, но остается неясным, в какой мере эксперты и группы неспециалистов способны адекватно распознавать и оценивать решения, основанные на внезапном и субъективном переживании инсайта. Этот пробел указывает на необходимость исследований, направленных на сопоставление экспертных и коллективных оценок, именно в контексте инсайтных решений.

Цель исследования

Настоящее исследование направлено на систематическую проверку возможности идентификации инсайтных решений по коротким видеофрагментам поведения участников в момент решения задач. Особое внимание уделялось вопросу согласованности оценок, полученных из разных источников. Анализ включал, во-первых, субъективные самоотчеты участников о переживании инсайта. Во-вторых, использовались экспертные оценки видеозаписей. Эксперты оценивали 90 видеофрагментов по 8 поведенческим проявлениям (позитивные и негативные эмоции, жестикуляция, мимика и т.д.) и определяли наличие/отсутствие инсайта. В-третьих, та же процедура была масштабирована на наивных оценщиков (студентов) через оценку всех 976 видеороликов с калибровкой на наборе эталонных клипов, согласованно оцененных экспертами. Основные вопросы, на которые мы искали ответ в настоящем исследовании, следующие.

1. Насколько согласовано эксперты оценивают инсайт и поведенческие проявления, с ним связанные?

2. Какова структура экспертных оценок поведенческих проявлений и можно ли по ним предсказать экспертные и субъективные оценки инсайтности?

3. Можно ли предсказать экспертные и субъективные оценки инсайтности с помощью агрегированных оценок поведения большим числом наивных респондентов?

4. Как связаны субъективные оценки инсайтности с экспертными оценками и оценками, предсказанными по оценкам наивных испытуемых? Различается ли связь в зависимости от типа переживаемого инсайта («Ага!» реакция / «Ах, да!» реакция)

Таким образом, работа проверяет возможность объективной видео-регистрации инсайта и анализирует потенциал привлечения широкого круга неспециалистов для задач оценки инсайта.

Описание основного эксперимента

Участники. В исследовании приняли участие 30 добровольцев (3 мужчины и 27 женщин) в возрасте от 22 до 54 лет ($M = 35,4$; $SD = 6,95$). Все участники были информированы о видеозаписи, подписали информированное согласие и принимали участие на безвозмездной основе.

Стимульный материал. В качестве стимульного материала использовались 20 задач на поиск отдаленных ассоциаций (RAT-RUS) из базы данных Н.В. Морошкиной с соавт. (Морошкина и др., 2021; Moroshkina et al., 2022).

Аппаратура и программное обеспечение. Эксперимент был реализован в среде PsychoPy 2023.2.3 и проводился индивидуально на ноутбуке HUAWEI MCLF-XX (разрешение экрана 1920×1200, 60 Гц). Видеозапись осуществлялась из двух источников: захват экрана и запись с веб-камеры, с использованием программы OBS Studio 30.1.2. Постобработка видеоматериалов производилась в программе Shotcut 24.04.28.

Процедура. Перед началом эксперимента все участники получали стандартную инструкцию и краткое описание феномена инсайта (Valueva, Lapteva, 2023). Решение задач проходило в 2 этапа¹. На решение каждой задачи на каждом этапе отводилось не более 60 с. Задачи предъявлялись участникам в случайном порядке. В случае верного решения участники отвечали на вопрос «Были ли у Вас инсайт?» с вариантами ответа Да/Нет). В случае неудачи на первом этапе задача переносилась на второй этап, а после неудачи на втором этапе демонстрировалось правильное решение, после

чего фиксировалось наличие/отсутствие «Ах, да!» реакции (переживание инсайта при предъявлении верного ответа). Перед началом основной серии испытуемые решали 2 тренировочные задачи.

Экспертная оценка видео: материалы и методы

Стимульный материал. В результате проведения эксперимента было получено 30 полноформатных видеопротоколов, фиксирующих процесс решения задач всеми участниками исследования. Все видеопротоколы были сегментированы на короткие видеоролики, охватывающие временной интервал непосредственно перед моментом, когда испытуемый давал ответ на задачу (или видел предъявленный ответ). Продолжительность одного видеоролика составляла 5–7 с.

В итоговый массив данных вошло 976 видеофрагментов. Для экспертной оценки был сформирован набор из 90 видеороликов — по три фрагмента от каждого участника, отражающих разные типы ситуаций: верное решение, неверное решение и участнику продемонстрировали ответ («Ах, да!» реакция).

Процедура. Для проведения оценки видеоматериала экспертами был разработан чат-бот в мессенджере Telegram. Видеофрагменты предъявлялись по одному, в случайном порядке, что позволяло минимизировать эффект последовательности. Оценка каждого видео осуществлялась по следующим поведенческим параметрам: выраженность положительных и отрицательных эмоций, внезапность решения, вербальные реакции, жестикуляция, двигательные проявления и мимика. Каждый параметр оценивался по 6-балльной шкале (от 0 — «очень слабо выражено» до 5 — «очень сильно выражено»). Дополнительно экспертам предлагался вопрос: «Был ли у решающего инсайт?» с вариантами ответа Да/Нет. Функционал чат-бота предусматривал автоматическое сохранение прогресса, что позволяло при необходимости приостанавливать и возобновлять процесс оценки без потери ранее внесенных данных.

Выборка. В исследовании приняли участие шесть экспертов в области изучения инсайта (три женщины и три мужчины), отобранных по критериям наличия значительного профессионального опыта и публикационной активности в данной предметной области.

Анализ выполнялся в программной среде R (version 4.5.1) с использованием библиотек *psych*

¹ Для целей другого исследования участники были случайным образом распределены в две группы (с инкубационным периодом и без него) и имели по 2 попытки на решение каждой задачи. Однако в целях настоящего исследования данные двух попыток были объединены и анализировались вместе.

(метод главных компонент), *lavaan* (конфирматорный факторный анализ), *irr* (каппа Флейса, коэффициент Кендалла), *rsq* (Nagelkerke псевдо- R^2), *pROC* (ROC-кривая и критерий Юдена). Логистическая регрессия оценивалась базовой функцией *glm*.

Экспертная оценка видео: результаты

Оценка экспертами инсайтности решений. Процентное соотношение выявленных инсайтных решений экспертами представлен на рис. 1.

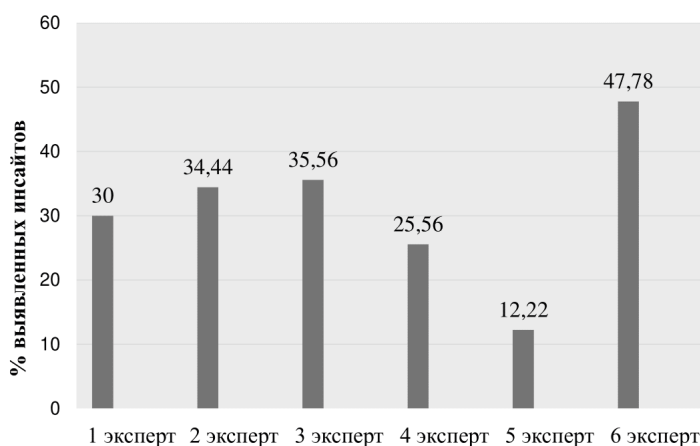


Рис. 1. Процент инсайтных решений, выявленных экспертами

Для количественной оценки степени согласованности между экспертами был рассчитан коэффициент каппа Флейса (Fleiss' Карра, K). Для всех шести экспертов значение составило $K=0,419$, что, согласно общепринятым критериям интерпретации данного коэффициента, соответствует умеренному уровню согласованности. Также был проведен дополнительный анализ межэксперт-

Диапазон между максимальной и минимальной оценками экспертов достиг 35.56 процентных пункта. Указанный диапазон оценок подчеркивает значительные различия во мнениях экспертов, что потенциально свидетельствует о недостаточной согласованности в оценке наличия инсайтов. Данные расхождения могут быть обусловлены субъективностью восприятия, различиями в интерпретации критериев инсайтности, а также индивидуальными различиями в подходах к оцениванию (ригидность/либеральность).

ной согласованности для подгруппы экспертов (№ 1, 2 и 3), выделенных на основании сходства в доле выявленных ими инсайтов (около 30%). В данной подгруппе значение $K=0,617$, что указывает на существенно более высокий уровень согласованности. Анализ согласованности экспертов отдельно по поведенческим параметрам представлен в табл. 1.

Таблица 1. Результаты коэффициент Кендалла (W): согласованность по поведенческим параметрам всех экспертов и подгруппы экспертов (1, 2, 3)

Поведенческий параметр	Согласованность по всем экспертам	Согласованность по трем экспертам
Положительные эмоции	0,507	0,594
Негативные эмоции	0,22	0,319
Удивление	0,51	0,632
Внезапность	0,392	0,612
Вербальные проявления	0,485	0,507
Жесты	0,432	0,547
Движения	0,383	0,544
Мимика	0,648	0,745

Оценка экспертами поведенческих проявлений. Коэффициент конкордации Кендалла (W) показал умеренную согласованность между экспер-

тами в оценке поведенческих параметров. Наиболее высокая согласованность наблюдалась для параметра «Мимика» как среди всех экспертов

($W=0,648$), так и в подгруппе из трех наиболее согласованных экспертов (1, 2 и 3) ($W=0,745$), что указывает на наибольшее единообразие в интерпретации мимических проявлений. Высокая согласованность также отмечалась для параметров «Удивление» и «Положительные эмоции», особенно в подгруппе из трех экспертов ($W=0,632$ и $0,594$ соответственно). Менее согласованными оказались оценки для «Негативных эмоций», «Внезапности» и «Движений», демонстрируя большую вариативность в интерпретации этих поведенческих проявлений среди экспертов. В целом, согласованность между экспертами была выше в подгруппе из трех наиболее согласованных экспертов по всем поведенческим параметрам, что подчеркивает наличие ядра экспертов с более схожим пониманием и оценкой наблюдаемых поведенческих сигналов. Таким образом, в дальнейшем мы будем опираться именно на подгруппу из данных трех экспертов.

Экспертные оценки (по 3 наиболее согласованным экспертам) по каждому из поведенческих проявлений были усреднены. На основании этих усредненных оценок был проведен эксплораторный факторный анализ (метод главных компонент). Результаты показали, что выделяется 2 фактора,

в совокупности объясняющих 68% дисперсии. На основе анализа 8 поведенческих параметров было выделено два латентных фактора. В Фактор 1 (43,6% дисперсии) вошли: «Позитивные эмоции» (нагрузка 0,996), «Мимика» (0,965), «Удивление» (0,991), «Внезапность» (0,843). В Фактор 2 — «Жестикуляция» (0,885), «Движения» (0,665) и «Негативные эмоции» (0,634). Переменная «Вербальные реакции» была нагружена сразу по двум факторам (0,272 и 0,549 соответственно) и была исключена из дальнейшего анализа, т.к. были свидетельства того, что в некоторых случаях звук у пользователей телеграм-бота не работал. Первый фактор был интерпретирован как «Позитивная экспрессия», отражающий общее восприятие экспертами поведения человека через аспект выразительности лица, отличный от общей интенсивности. Второй фактор, имеющий высокие нагрузки по переменным, связанным с жестами, движениями и негативными эмоциями, был назван нами «Негативная экспрессия» и может быть интерпретирован как общая моторная напряженность. Конфирматорный факторный анализ подтвердил двухфакторную структуру оценки поведенческих проявлений экспертами ($\chi^2(12)=22,06$; $p=0,037$; CFI=0,98; TLI=0,96; RMSEA=0,097; SRMR=0,036) (рис. 2).

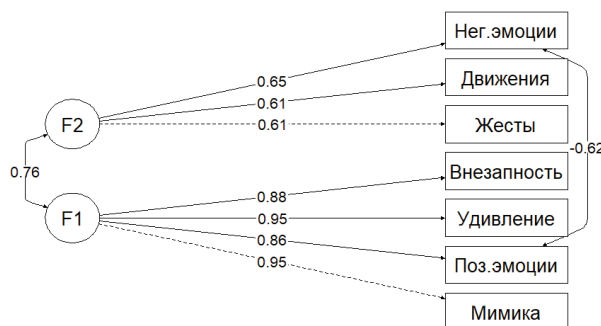


Рис. 2. Факторная модель оценки поведенческих проявлений экспертами

Предсказание инсайтности на основе поведенческих проявлений. Экспертные оценки инсайтности решений были переведены в шкалу 0/1 по принципу «большинства» (т.е. если из 3 экспертов двое считали, что инсайт был, то решение признавалось инсайтным, и наоборот — если 2 из трех считали, что инсайта не было, решение признавалась неинсайтным). Был проведен логистический регрессионный анализ, в котором зависимой переменной выступили полученные оценки инсайтности, а независимыми — 2 фактора (По-

зитивная экспрессия и Негативная экспрессия). Выявлено значимое влияние фактора Позитивная экспрессия ($\beta=4,46$; $p<0,001$), а влияние фактора Негативная экспрессия оказалось не значимым ($\beta=0,85$; $p=0,167$). Nagelkerke's R^2 (процент объясняемой дисперсии модели)² составил 0,76, что является достаточно высоким показателем.

На следующем шаге мы отобрали только те видеофрагменты ($n=67$), по которым согласованность трех экспертов являлась стопроцентной (т.е. все три эксперта признавали решение инсайт-

² Nagelkerke's R^2 — это одна из мер псевдо- R^2 , с помощью которой часто оценивается качество подгонки модели (UCLA Statistical Consulting Group, n.d.). Коэффициент Нагелкерке R^2 масштабирован так, чтобы он мог принимать значения от 0 до 1, поэтому обычно интуитивно трактуется как доля объясненной дисперсии.

ным или неинсайтным). Регрессионный анализ, проведенный для этих видео, продемонстрировал сходные результаты: значимое влияние фактора Позитивная экспрессия ($\beta=7,78$; $p=0,038$), и незначимое влияние фактора Негативная экспрессия ($\beta=1,90$; $p=0,261$). Nagelkerke's R^2 , однако, составил 0,92, что является очень высоким показателем.

Такой же анализ был проведен для предсказания субъективных оценок инсайтности (т.е. оценок самими испытуемыми способа решения задачи) по экспертным оценкам поведенческих проявлений. Из 90 оцененных экспертами видеофрагментов в анализ вошел 61, по которому имелись субъективные оценки инсайтности. Результаты регрессионного анализа показали, факторы Позитивной и Негативной экспрессии не являются значимыми предикторами оценок инсайтности, а показатель Nagelkerke's R^2 составил 0,13, что является низким значением.

Таким образом, эксперты при оценке инсайтности решений в основном опираются на позитивные эмоциональные проявления, игнорируя двигательную активность и негативные эмоции. По экспертным оценкам поведенческих проявлений предсказать с достаточной степенью надежности субъективные оценки инсайтности невозможно.

Предсказание инсайтности решений на основании оценок наивных испытуемых

Выборка. В исследовании приняли участие 283 человека, включая 130 мужчин и 140 женщин в возрасте от 18 до 53 лет ($M=21,04$; $SD=4,53$), 13 участников не указали пол и возраст. Основную часть выборки составили студенты ФГАОУ ВО «Казанский (Приволжский) федеральный университет», а также добровольцы, привлеченные по принципу «снежного кома». Часть участников была приглашена через личные контакты и рекомендации знакомых. Испытуемые случайным образом были разбиты на 21 группу, каждая группа испытуемых оценивала уникальный видеосет.

Стимульный материал. Каждый видеосет включал в себя: 1) 37 эталонных (одинаковых для всех) видеофрагментов, согласованно оцененных экспертами — это служило «калибровкой» для наивных оценщиков; 2) 42 либо 43 новых видеофрагментов из исходной коллекции. Эти видео были «тестовыми» неразмеченными фрагментами, для которых нужно было получить оценки инсайтности. Таким образом, каждый испытуемый должен был оценить 79 или 80 видео. Такой дизайн позволил сделать охват всех видеофрагментов, избегая непосильной нагрузки на одного испытуемого.

Процедура исследования. Испытуемые оценивали видео в том же Telegram-боте и по той же процедуре, что и эксперты, за исключением того, что им не предъявлялся вопрос «Был ли у испытуемого инсайт?».

Предварительный анализ данных. Из 283 испытуемых, зарегистрировавшихся Telegram-боте, 92 человека либо не закончили оценку видео (оценили менее 70 видео), либо имели нулевой разброс оценок. Таким образом, в анализе остались данные 191 испытуемого, на каждый сет приходилось от 6 до 12 человек ($M=9,09$; $SD=1,64$).

Факторный анализ. Анализ проводился по каждому сету отдельно. Оценки по всем шкалам, кроме шкалы Негативные эмоции, были подвергнуты анализу методом главных компонент с выделением одного генерального фактора. Шкала негативных эмоций для большинства видеосетов очень слабо (или отрицательно) коррелировала с остальными шкалами, поэтому анализировалась отдельно. Индекс КМО варьировал от 0,57 до 0,88 ($M=0,78$), тест Бартлетта во всех датасетах был значимым. Процент объясненной дисперсии для первого фактора составил от 0,28 до 0,67. Для каждого видеосета была получена агрегированная оценка испытуемых по всем шкалам за исключением негативных эмоций, представляющая собой линейную комбинацию стандартизированных переменных с факторными весами.

Регрессионный анализ: прогноз экспертных оценок. На первом шаге были использованы данные по 37 эталонным видео, для которых были получены однозначные экспертные оценки по наличию/отсутствию инсайта (обучающая выборка). Для каждого видеосета была проведена отдельная логистическая регрессия, в которой зависимой переменной было наличие/отсутствие инсайта, а предикторами — факторные нагрузки, полученные в результате факторизации на предыдущем шаге, и отдельно — негативные эмоции.

Оценивание параметров выполнялось методом максимального правдоподобия. Качество подгонки оценивалось с помощью Nagelkerke's R^2 . Для выбора порога классификации в каждом видеосете использовался индекс Юдена ($J = \text{sensitivity} + \text{specificity} - 1$), с выбором порога, максимизирующего J на ROC-кривой. Для 21 видеосета Nagelkerke's R^2 варьировал от 0,15 до 0,94 ($M=0,61$), оптимальные пороги — от 0,23 до 0,68.

На втором шаге, используя полученные значения интерсептов и регрессионных коэффициентов, мы получили оценки вероятностей для тех видео, у которых экспертные оценки отсутствовали (по формуле $\text{logit}(p_{\text{insight}}) = b_0 + b_1 \times \text{ФакторПоз}$).

Эмоций + $b_2 \times \text{Нег.Эмоции}$). Классификация осуществлялась на основании оптимального порога, вычисленного на эталонных видео для каждого видеосета.

В результате для каждого видео были получены оценки вероятностей наличия или отсутствия инсайта, а также классификация 0/1 на основании оптимальных пороговых значений.

Регрессионный анализ: прогноз субъективных оценок. Дополнительно был проведен анализ с целью выяснить, насколько хорошо можно предсказать субъективные оценки инсайтности по оценкам поведенческих проявлений наивных испытуемых. Анализ проводился для каждого сета отдельно. Результаты показали, что субъективные оценки предсказываются гораздо хуже экспертных (Nagelkerke's R^2 варьировал от 0 до 0,3; $M=0,14$)

и только в 9 видеосетах из 21 фактор Позитивных Эмоций оказался значимым предиктором.

Связь субъективных оценок, экспертных оценок и оценок наивных испытуемых

Для проверки согласованности разных типов оценок инсайтности (экспертных, субъективных и предсказанных на основе поведенческих оценок наивных испытуемых) был проведен корреляционный анализ (см. табл. 2). Результаты демонстрируют, что сила связи варьирует в зависимости как от типа инсайтной реакции («Ага!» или «Ах, да!»), так и от опыта испытуемого (эксперт или наивный участник). В целом, экспертные оценки умеренно коррелируют с субъективными, причем наиболее выраженная связь наблюдается для реакций, характеризующихся подлинным озарением («Ага!»).

Тип инсайта	Эксперты	Наивные испытуемые
В целом	0,34 (p=0,007) n=61	0,18 (p<0,001) N=529
Ага! – реакция	0,47 (p=0,006) n=32	0,16 (p=0,005) n=299
Ах, да! – реакция	0,15 (p=0,429) n=29	0,20 (p=0,002) n=230

Обсуждение результатов

Полученные данные показывают, что экспертная идентификация инсайтного решения по коротким видеозаписям поведения возможна, но чувствительна к критериям и стратегиям оценивания. Общая межэкспертная согласованность оказалась умеренной ($K=0,419$), при этом внутри подгруппы трех экспертов, сходных по доле выявляемых инсайтов, согласованность была заметно выше ($K=0,617$). Такое расслоение показывает, что среди специалистов существуют различия в понимании инсайта и в порогах отнесения решения к инсайтному. С практической точки зрения это означает, что использование экспертных оценок требует либо предварительной калибровки и обучения оценщиков, либо применения процедур консенсуса и контроля «строгости» эксперта.

Анализ согласованности по отдельным поведенческим шкалам продемонстрировал, что наиболее согласованно эксперты оценивали мимику, а также удивление и позитивные эмоции. Напротив, оценки двигательной активности, внезапности и негативных эмоций имели большую вариативность. Таким образом, можно сделать вывод, что лицевые экспрессивные проявления и пози-

тивный аффект являются наиболее доступными для внешнего наблюдения и наиболее однозначно интерпретируемыми поведенческими признаками. Полученная факторная структура экспертных оценок поведения свидетельствует о большей степени дифференцированности суждений экспертов по сравнению с наивными испытуемыми. Экспертные оценки поведенческих проявлений продемонстрировали четкую двухфакторную структуру, в то время как у студентов все оценки, за исключением шкалы негативных эмоций, в большинстве случаев сливались в один фактор, который можно интерпретировать как общую активность (см. также: Кулиев, Валуева, Куприянов, 2026). Основным предиктором решения о наличии инсайта у экспертов выступал фактор позитивной экспрессии, тогда как у наивных испытуемых именно фактор общей активности лучше всего предсказывал экспертные оценки инсайтности. В целом эти результаты согласуются с данными о соматических и экспрессивных маркерах озарения (Shen et al., 2018; Laukkonen et al., 2021). В обоих случаях негативная экспрессия практически не вносила вклада в процент объясняемой дисперсии. При этом предсказание экспертных

оценок на основе поведенческих оценок наивных испытуемых можно считать в среднем достаточно успешным (средний Nagelkerke's $R^2=0,61$ по 21 видеосету). Даже без специальной подготовки неспециалисты способны, при достаточной статистической агрегации, воспроизводить экспертные оценки. С одной стороны, этот факт открывает перспективы краудсорсинговой разметки больших массивов видеоданных — при наличии небольшого набора «якорных» примеров можно получать приближенные к экспертным оценки наличия инсайта на больших массивах видеоданных с существенно меньшими ресурсными затратами.

С другой стороны, данные свидетельствуют о том, что экспертная оценка в большей степени чувствительна к качественным различиям в паттернах поведения, тогда как наивная оценка отражает скорее общую заметность реакции, чем дифференцированное распознавание специфических маркеров инсайта. Следовательно, при использовании наивных испытуемых в качестве оценщиков желательна предварительная калибровка критериев и процедуры обучения, позволяющие сместить их внимание от общей поведенческой активности к более специфическим признакам инсайтного решения.

Высокая предсказуемость экспертной классификации по поведенческим оценкам не тождественна высокой валидности «внешнего» критерия инсайта. Скорее, результаты указывают на набор признаков, которыми пользуются эксперты для принятия решения об инсайте. Такой набор может быть эффективен для детекции ярких «Ага!» решений, но потенциально хуже работает для инсайтов без выраженной экспрессии, а также для ситуаций, когда позитивная эмоциональная реакция связана не с переструктурированием, а с простым облегчением после правильного ответа.

Проблема соотношения объективных и субъективных критериев инсайтности в нашем исследовании проявилась в виде слабой связи между наблюдаемыми поведенческими проявлениями и субъективными самоотчетами участников. Ни экспертные оценки поведения, ни агрегированные оценки наивных наблюдателей не позволили надежно предсказать субъективную инсайтность. При этом корреляция субъективных оценок с экспертными также оказалась не слишком высокой и заметно зависела от типа инсайтной реакции: для «Ага!» связь была выше, чем для «Ах, да!» реакции. Этот результат свидетельствует о выраженной диссоциации между наблюдаемым поведением и самоотчетом, что согласуется с данными о том, что решатели нередко ретроспективно ре-

конструируют процесс решения или недооценивают внезапность озарения, а вербализация может искажать доступ к неявным компонентам процесса решения (Schooler et al., 1993; Danek et al., 2020).

Представленная работа имеет ряд ограничений. Во-первых, следует отметить сравнительно невысокую согласованность между экспертами при определении наличия или отсутствия инсайта у решателя. Это ограничение указывает не только на сложность самой процедуры оценивания, но и на то, что даже среди специалистов сохраняются различия в понимании инсайта и в критериях его распознавания. Экспертные расхождения в данном случае отражают не просто индивидуальную вариативность суждений, но и более общую концептуальную неоднородность в трактовке того, какие признаки следует считать характерными для инсайтного решения. В этом смысле экспертная оценка выступает не как полностью однозначный критерий, а как операционализация одной из возможных интерпретаций инсайта, что необходимо учитывать при обсуждении полученных результатов.

Во-вторых, несмотря на в среднем хороший результат масштабирования экспертных оценок с помощью оценок наивных испытуемых, мы наблюдали достаточно высокую вариативность между видеосетами, что указывает на необходимость процедур контроля качества: отсев участников с низкой вовлеченностью, уточнение шкал, обучающие процедуры, а также проверка устойчивости моделей при смене состава оценивающих групп.

В-третьих, анализ проводился по коротким видеозаписям, что, с одной стороны, позволяет стандартизировать процедуру оценивания, но, с другой стороны, ограничивает доступ наблюдателей к более широкому контексту решения. Часть поведенческих признаков может проявляться до или после выделенного фрагмента, а их интерпретация может зависеть от хода решения в целом. Соответственно, часть случаев инсайта, особенно менее экспрессивных, могла остаться незамеченной.

В-четвертых, в качестве «объективного» критерия в работе использовалась экспертная классификация. Успешное воспроизведение экспертных оценок не тождественно универсальной валидности предложенного подхода и требует дальнейшей проверки на других типах инсайтных задач, других форматах поведенческого материала и при использовании независимых объективных индикаторов.

Заключение

Проведенное исследование показало, что инсайтные решения в принципе могут распознаваться по коротким видеозаписям поведения, однако

надежность такой идентификации зависит от того, кто именно оценивает материал и какими критериями он руководствуется. Экспертные оценки продемонстрировали умеренную общую согласованность и более высокую согласованность в более однородной подгруппе, что указывает на наличие различий в трактовке инсайта даже среди специалистов. Структура экспертных оценок оказалась более дифференцированной, чем у наивных испытуемых, а ключевым предиктором экспертного решения о наличии инсайта выступила позитивная экспрессия. Агрегированные оценки наивных наблюдателей в среднем позволяли достаточно успешно приближаться к экс-

пертным оценкам, что подтверждает потенциал калиброванной коллективной разметки. При этом ни экспертные, ни агрегированные поведенческие оценки не обеспечили надежного прогноза субъективной инсайтности, а связь с самоотчетом оказалась ограниченной и более выраженной для реакций типа «Ага!», чем для «Ах, да!». Тем самым исследование подтверждает перспективность анализа видео как инструмента объективной регистрации внешних проявлений инсайта, но одновременно показывает, что внешние поведенческие признаки и субъективное переживание инсайта совпадают лишь частично.

Литература:

- Владимиров И.Ю., Бушманова А.С., Курицын А.А. Улыбка — залог успеха: анализ поведенческих маркеров инсайтного решения // Психология познания: материалы Всероссийской научной конференции. ЯрГУ, 16–17 декабря 2022 г. /отв. ред. И.Ю. Владимиров, С.Ю. Коровкин. Ярославль: Филигрань, 2023. С. 56–60.
- Владимиров И.Ю., Бушманова А.С., Макаров И.Н. Поведенческие маркеры ключевых событий инсайтного решения // Психологические исследования. 2021. Т. 14. № 77. DOI: 10.54359/ps.v14i77.134
- Гаршина С.А., Пискарев П.М. Анализ современных методов детекции инсайтности // Современные технологии управления. 2023. № 3(103). С. 1. DOI: 10.24412/2226-9339-2023-3103-1
- Кулиев Н.С., Валуева Е.А., Куприянов Р.В. Можно ли увидеть инсайт? Сравнение экспертных и наивных оценок поведенческих признаков // Психология творчества: традиции, инновации, перспективы. Материалы Международной научной конференции, посвященной 105-летию со дня рождения Я.А. Пономарева / отв. ред. С.Ю. Коровкин, Д.В. Ушаков, А.Л. Журавлев. М.: Изд-во «Институт психологии РАН», 2025. С. 66–69.
- Лазарева Н.Ю., Савинова А.Д., Чистопольская А.В. Влияние экспериментальных условий на субъективную оценку инсайтности решения // Экспериментальная психология. 2023. Т. 16, № 1. С. 23–42. DOI: 10.17759/exppsy.2023160102
- Морошкина Н.В., Аммайнен А.В., Гершкович В.А., Львова О.В., Федосова В.И. Апробация русскоязычного банка заданий на поиск отдаленных ассоциаций (RAT-Rus) для исследований инсайта // Первый национальный конгресс по когнитивным исследованиям, искусственному интеллекту и нейроинформатике. Девятая международная конференция по когнитивной науке: сборник научных трудов. Т. 1. М.: МИФИ, 2021. С. 559–561.
- Морошкина Н.В., Аммайнен А.В., Савина А.В. В погоне за инсайтом: современные подходы и методы измерения инсайта в когнитивной психологии // Психологические исследования. 2020. Т. 13. № 74. DOI: 10.54359/ps.v13i74.163
- Филяева О.В., Коровкин С.Ю. Объективные маркеры инсайтного решения // Творчество: наука, искусство, жизнь. 2015. С. 367–371.
- Чистопольская А.В., Савинова А.Д., Лазарева Н.Ю. Экспликация критериев инсайта и обзор методов их измерения // Психология. Журнал Высшей школы экономики. 2021. Т. 18, № 4. С. 907–929. DOI: 10.17323/1813-8918-2021-4-907-929
- Шумилов Т.В., Чистопольская А.В. Уровень развития метакогнитивных способностей как коррелят точности детекции инсайтного решения // Психология познания: материалы Всероссийской научной конференции. ЯрГУ, 16–17 декабря 2022 г. /отв. ред. И.Ю. Владимиров, С.Ю. Коровкин. Ярославль: Филигрань, 2023. С. 361–366.
- Adler E. Modeling the Affect of “Aha!” moments to detect the moment of learning: doctoral dissertation. Cambridge (MA): Massachusetts Institute of Technology. 2024.
- Bowden E.M., Jung-Beeman M. Aha! Insight experience correlates with solution activation in the right hemisphere // Psychonomic Bulletin & Review. 2003. Vol. 10(3). P. 730–737. DOI: 10.3758/BF03196539
- Cernea D., Kerren A., Ebert A. Detecting insight and emotion in visualization applications with a commercial EEG headset // SIGRAD 2011 Conference Proceedings. 2011. P. 53–60.

- Chesebrough C., Oh Y., Kounios J. Why my “Aha!” is your “Hmm...”: Individual differences in the phenomenology and likelihood of insight experiences // *The Emergence of Insight*. Cambridge: Cambridge University Press. 2024. P. 251–279. DOI: 10.1017/9781009244244.012
- Cropley D.H., Kaufman J.C. Measuring functional creativity: Non-expert raters and the Creative Solution Diagnosis Scale // *The Journal of Creative Behavior*. 2012. Vol. 46(2). P. 119–137. DOI: 10.1002/jocb.9
- Danek A.H. Magic tricks, sudden restructuring, and the Aha! experience: A new model of nonmonotonic problem solving // *Insight: On the Origins of New Ideas* / ed. by F. Vallee-Tourangeau. London; New York: Routledge. 2018. P. 51–78. DOI: 10.4324/9781315268118-4
- Danek A.H., Fraps T., von Muller A., Grothe B., Ollinger M. It’s a kind of magic: what self-reports can reveal about the phenomenology of insight problem solving // *Frontiers in Psychology*. 2014. Vol. 5. Art. 1408. DOI: 10.3389/fpsyg.2014.01408
- Danek A.H., Wiley J. What about false insights? Deconstructing the Aha! experience along its multiple dimensions for correct and incorrect solutions separately // *Frontiers in Psychology*. 2017. Vol. 7. Art. 2077. DOI: 10.3389/fpsyg.2016.02077
- Danek A.H., Williams J., Wiley J. Closing the gap: Connecting sudden representational change to the subjective Aha! experience in insightful problem solving // *Psychological Research*. 2020. Vol. 84. Vol. 1. P. 111–119. DOI: 10.1007/s00426-018-0977-8
- Galton F. Vox Populi // *Nature*. 1907. Vol. 75. P. 450–451. DOI: 10.1038/075450a0
- Ishikawa T., Toshima M., Mogi K. How and when? Metacognition and solution timing characterize an “aha” experience of object recognition in hidden figures // *Frontiers in Psychology*. 2019. Vol. 10. Art. 1023. DOI: 10.3389/fpsyg.2019.01023
- Kameda T., Toyokawa W., Tindale R.S. Information aggregation and collective intelligence beyond the wisdom of crowds // *Nature Reviews Psychology*. 2022. Vol. 1(6). P. 345–357. DOI: 10.1038/s44159-022-00054-y
- Kaufman J.C., Baer J., Cole J.C. Expertise, domains, and the consensual assessment technique // *The Journal of Creative Behavior*. 2009. Vol. 43(4). P. 223–233. DOI: 10.1002/j.2162-6057.2009.tb01316.x
- Kaufman J.C., Baer J., Cole J.C., Sexton J.D. A comparison of expert and nonexpert raters using the consensual assessment technique // *Creativity Research Journal*. 2008. Vol. 20(2). P. 171–178. DOI:10.1080/10400410802059929
- Kaufman J.C., Baer J., Cropley D.H., Reiter-Palmon R., Sinnott S. Furious activity vs. understanding: How much expertise is needed to evaluate creative work? // *Psychology of Aesthetics, Creativity, and the Arts*. 2013. Vol. 7(4). P. 332–340. DOI: 10.1037/a0034809
- Knoblich G., Ohlsson S., Raney G.E. An eye movement study of insight problem solving // *Memory & Cognition*. 2001. Vol. 29(7). P. 1000–1009. DOI: 10.3758/BF03195762
- Korovkin S., Vladimirov I., Chistopolskaya A., Savinova A. How working memory provides representational change during insight problem solving // *Frontiers in Psychology*. 2018. Vol. 9. Art. 1864. DOI: 10.3389/fpsyg.2018.01864
- Kounios J., Beeman M. The cognitive neuroscience of insight // *Annual Review of Psychology*. 2014. Vol. 65. P. 71–93. DOI: 10.1146/annurev-psych-010213-115154
- Laukkonen R.E., Ingledew D.J., Grimmer H.J., Schooler J.W., Tangen J.M. Getting a grip on insight: Real-time and embodied Aha experiences predict correct solutions // *Cognition and Emotion*. 2021. Vol. 35(5). P. 918–935. DOI: 10.1080/02699931.2021.1908230
- Laukkonen R.E., Tangen J.M. How to detect insight moments in problem solving experiments // *Frontiers in Psychology*. 2018. Vol. 9. Art. 282. DOI: 10.3389/fpsyg.2018.00282
- Moroshkina N.V., Savina A.I., Ammalainen A.V., Gershkovich V.A., Zverev I.V., Lvova O.V. How difficult was it? Metacognitive judgments about problems and their solutions after the Aha moment // *Frontiers in Psychology*. 2022. Vol. 13. Art. 911904. DOI: 10.3389/fpsyg.2022.911904
- Salvi C., Bricolo E., Franconeri S.L., Kounios J., Beeman M. Sudden insight is associated with shutting out visual inputs // *Psychonomic Bulletin & Review*. 2015. Vol. 22(6). P. 1814–1819. DOI: 10.3758/s13423-015-0845-0
- Salvi C., Bricolo E., Kounios J., Bowden E.M., Beeman M. Insight solutions are correct more often than analytic solutions // *Thinking & Reasoning*. 2016. Vol. 22(4). P. 443–460. DOI: 10.1080/13546783.2016.1141798
- Salvi C., Simoncini C., Grafman J., Beeman M. Oculometric signature of switch into awareness? Pupil size predicts sudden insight whereas microsaccades predict problem-solving via analysis // *NeuroImage*. 2020. Vol. 217. Art. 116933. DOI: 10.1016/j.neuroimage.2020.116933
- Schooler J.W., Ohlsson S., Brooks K. Thoughts beyond words: When language overshadows insight // *Journal of Experimental Psychology: General*. 1993. Vol. 122(2). P. 166–183. DOI: 10.1037/0096-3445.122.2.166

- Shen W., Tong Y., Yuan Y., Zhan H., Liu C., Luo J., Cai H. Feeling the insight: Uncovering somatic markers of the “aha” experience // *Applied Psychophysiology and Biofeedback*. 2018. Vol. 43(1). P. 13–21. DOI: 10.1007/s10484-017-9381-1
- Shumilov T.V., Chistopolskaya A.V., Vladimirov I.Y. The Path to Insight: Developing Methods to Increase the Effectiveness of Insightful Solution Detection // *Psychology. Journal of Higher School of Economics*. 2023. Vol. 20(3). P. 473–489. DOI: 10.17323/1813-8918-2023-3-473-489
- Spiridonov V., Loginov N., Ardislamov V. Dissociation between the subjective experience of insight and performance in the CRA paradigm // *Journal of Cognitive Psychology*. 2021. Vol. 33(6–7). P. 685–699. DOI: 10.1080/20445911.2021.1900198
- Storme M., Myszkowski N., Celik P., Lubart T. Learning to judge creativity: The underlying mechanisms in creativity training for non-expert judges // *Learning and Individual Differences*. 2014. Vol. 32. P. 19–25. DOI: 10.1016/j.lindif.2014.03.002
- Tsai K.C. Quasi-experts’ assessments of creative products: An evaluation using a sample of design students // *British Journal of Education, Society & Behavioural Science*. 2016. Vol. 12(3). P. 1–7. DOI: 10.9734/BJESBS/2016/21057
- Valueva E.A., Lapteva N.M. “Aha!” and “Oh Yes!”: How emotions affect insight experience // *Psychology. Journal of Higher School of Economics*. 2023. Vol. 20(3). P. 428–444. DOI: 10.17323/1813-8918-2023-3-428-444

CAPTURING INSIGHT ON VIDEO: EXPERT AND NAIVE EVALUATIONS OF BEHAVIORAL MARKERS IN INSIGHTFUL PROBLEM SOLVING

© Nikolay S. Kuliev

PhD student, Institute of Psychology of the Russian Academy of Sciences, Moscow, Russia
kuliev.nikolaj@bk.ru, ORCID: 0009-0004-1003-464X

© Ekaterina A. Valueva

PhD in Psychology, Institute of Psychology of the Russian Academy of Sciences, Moscow, Russia
valuevaea@ipran.ru, ORCID: 0000-0003-3637-287X

© Roman V. Kupriyanov

PhD in Psychology, Kazan Federal University, Kazan, Russia
rvkupriyanov@kpfu.ru, ORCID: 0000-0001-9794-9607

This study was supported by the Russian Science Foundation (RSF), project № 23–18–01059-P

The article checks the possibility of objective video registration of insight solutions using short video clips of solvers’ behavior. Experts evaluated 90 videos on several behavioral scales and classified solutions as insightful or non-insightful. The same approach was then scaled to 976 video clips with naive raters and calibration on reference examples. It is shown that inter-expert agreement in identifying insight is moderate, but it increases substantially in a more consistent subgroup of experts. Facial expressions, surprise, and positive emotions are identified most reliably. Expert ratings of behavioral signs form a two-factor structure, in which the decision about the presence of insight is mainly determined by the factor of positive expression. Aggregated ratings of naive participants reproduce expert decisions quite well on average, but they rely more on the overall salience of the reaction than on differentiated recognition of specific insight markers. At the same time, the connection between both expert and aggregated behavioral ratings with the solvers’ subjective self-reports remains weak. The results show that video analysis and calibrated collective labeling are promising for insight research, but they also underline the need for careful interpretation of external behavioral criteria.

Keywords: insight, aha-experience, self-report, video analysis, behavioral markers, expert ratings, naive raters, collective rating, crowdsourcing

REFERENCES

- Vladimirov I. Yu., Bushmanova A. S., Kuritsyn A. A. (2023). Smile is the key to success: Analysis of behavioral markers of insight solution // *Psychology of Cognition: Proceedings of the All-Russian Scientific Conference*. Yaroslavl State University, December 16–17, 2022. Ed. by I. Yu. Vladimirov, S. Yu. Korovkin. Yaroslavl: Filigran Publ. P. 56–60.
- Vladimirov I. Yu., Bushmanova A. S., Makarov I. N. (2021). Behavioral markers of key events in insight solution // *Psychological Studies*. Vol. 14(77). DOI: 10.54359/ps.v14i77.134
- Garshina S. A., Piskarev P. M. (2023). Analysis of modern methods for detecting insightfulness // *Modern Management Technology*. Vol. 3(103). P. 1. DOI: 10.24412/2226-9339-2023-3103-1
- Kuliev N.S., Valueva E.A., Kupriyanov R.V. (2025). Can insight be seen? A comparison of expert and naive ratings of behavioral signs // *Psychology of Creativity: Traditions, Innovations, Prospects. Proceedings of the International Scientific Conference Dedicated to the 105th Anniversary of Ya.A. Ponomarev*. Ed. by S. Yu. Korovkin, D. V. Ushakov, A.L. Zhuravlev. Moscow: Institute of Psychology RAS Publ. P. 66–69.
- Lazareva N.Yu., Savinova A.D., Chistopolskaya A.V. (2023). The influence of experimental conditions on the subjective evaluation of solution insightfulness // *Experimental Psychology*. Vol. 16(1). P. 23–42. DOI: 10.17759/exppsy.2023160102
- Moroshkina N.V., Ammalainen A.V., Gershkovich V.A., Lvova O.V., Fedosova V.I. (2021). Validation of the Russian-language Remote Associates Test item bank (RAT-Rus) for insight research // *The First National Congress on Cognitive Research, Artificial Intelligence and Neuroinformatics. The Ninth International Conference on Cognitive Science: Proceedings*. Vol. 1. Moscow: MIFI Publ. P. 559–561.
- Moroshkina N.V., Ammalainen A.V., Savina A.V. (2020). In pursuit of insight: Modern approaches and methods for measuring insight in cognitive psychology // *Psychological Studies*. Vol. 13(74). DOI: 10.54359/ps.v13i74.163
- Filyaeva O.V., Korovkin S. Yu. (2015). Objective markers of insight solution // *Creativity: Science, Art, Life*. P. 367–371.
- Chistopolskaya A.V., Savinova A.D., Lazareva N. Yu. (2021). Explication of insight criteria and a review of methods for their measurement // *Psychology. Journal of the Higher School of Economics*. Vol. 18(4). P. 907–929. DOI: 10.17323/1813-8918-2021-4-907-929
- Shumilov T.V., Chistopolskaya A.V. (2023). The level of metacognitive abilities as a correlate of the accuracy of insight solution detection // *Psychology of Cognition: Proceedings of the All-Russian Scientific Conference*. Yaroslavl State University, December 16–17, 2022. Ed. by I. Yu. Vladimirov, S. Yu. Korovkin. Yaroslavl: Filigran Publ. P. 361–366.
- Adler E. (2024). Modeling the Affect of “Aha!” moments to detect the moment of learning: doctoral dissertation. Cambridge (MA): Massachusetts Institute of Technology.
- Bowden E.M., Jung-Beeman M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere // *Psychonomic Bulletin & Review*. Vol. 10. Vol. 3. P. 730–737. DOI: 10.3758/BF03196539
- Cernea D., Kerren A., Ebert A. (2011). Detecting insight and emotion in visualization applications with a commercial EEG headset // *SIGRAD 2011 Conference Proceedings*. P. 53–60.
- Chesebrough C., Oh Y., Kounios J. (2024). Why my “Aha!” is your “Hmm...”: Individual differences in the phenomenology and likelihood of insight experiences // *The Emergence of Insight*. Cambridge: Cambridge University Press. P. 251–279. DOI: 10.1017/9781009244244.012
- Cropley D.H., Kaufman J.C. (2012). Measuring functional creativity: Non-expert raters and the Creative Solution Diagnosis Scale // *The Journal of Creative Behavior*. Vol. 46(2). P. 119–137. DOI: 10.1002/job.9
- Danek A.H. (2018). Magic tricks, sudden restructuring, and the Aha! experience: A new model of nonmonotonic problem solving // *Insight: On the Origins of New Ideas* / ed. by F. Vallee-Tourangeau. London; New York: Routledge. P. 51–78. DOI: 10.4324/9781315268118-4
- Danek A.H., Fraps T., von Muller A., Grothe B., Ollinger M. (2014). It’s a kind of magic: what self-reports can reveal about the phenomenology of insight problem solving // *Frontiers in Psychology*. Vol. 5. Art. 1408. DOI: 10.3389/fpsyg.2014.01408
- Danek A.H., Wiley J. (2017). What about false insights? Deconstructing the Aha! experience along its multiple dimensions for correct and incorrect solutions separately // *Frontiers in Psychology*. Vol. 7. Art. 2077. DOI: 10.3389/fpsyg.2016.02077
- Danek A.H., Williams J., Wiley J. (2020). Closing the gap: Connecting sudden representational change to the subjective Aha! experience in insightful problem solving // *Psychological Research*. Vol. 84. Vol. 1. P. 111–119. DOI: 10.1007/s00426-018-0977-8

- Galton F. (1907). *Vox Populi* // *Nature*. Vol. 75. P. 450–451. DOI: 10.1038/075450a0
- Ishikawa T., Toshima M., Mogi K. (2019). How and when? Metacognition and solution timing characterize an “aha” experience of object recognition in hidden figures // *Frontiers in Psychology*. Vol. 10. Art. 1023. DOI: 10.3389/fpsyg.2019.01023
- Kameda T., Toyokawa W., Tindale R. S. (2022). Information aggregation and collective intelligence beyond the wisdom of crowds // *Nature Reviews Psychology*. Vol. 1(6). P. 345–357. DOI: 10.1038/s44159-022-00054-y
- Kaufman J. C., Baer J., Cole J. C. (2009). Expertise, domains, and the consensual assessment technique // *The Journal of Creative Behavior*. Vol. 43(4). P. 223–233. DOI: 10.1002/j.2162-6057.2009.tb01316.x
- Kaufman J. C., Baer J., Cole J. C., Sexton J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique // *Creativity Research Journal*. Vol. 20(2). P. 171–178. DOI:10.1080/10400410802059929
- Kaufman J. C., Baer J., Copley D. H., Reiter-Palmon R., Sinnott S. (2013). Furious activity vs. understanding: How much expertise is needed to evaluate creative work? // *Psychology of Aesthetics, Creativity, and the Arts*. Vol. 7(4). P. 332–340. DOI: 10.1037/a0034809
- Knoblich G., Ohlsson S., Raney G. E. (2001). An eye movement study of insight problem solving // *Memory & Cognition*. Vol. 29(7). P. 1000–1009. DOI: 10.3758/BF03195762
- Korovkin S., Vladimirov I., Chistopolskaya A., Savinova A. (2018). How working memory provides representational change during insight problem solving // *Frontiers in Psychology*. Vol. 9. Art. 1864. DOI: 10.3389/fpsyg.2018.01864
- Kounios J., Beeman M. (2014). The cognitive neuroscience of insight // *Annual Review of Psychology*. Vol. 65. P. 71–93. DOI: 10.1146/annurev-psych-010213-115154
- Laukkonen R. E., Ingledew D. J., Grimmer H. J., Schooler J. W., Tangen J. M. (2021). Getting a grip on insight: Real-time and embodied Aha experiences predict correct solutions // *Cognition and Emotion*. Vol. 35(5). P. 918–935. DOI: 10.1080/02699931.2021.1908230
- Laukkonen R. E., Tangen J. M. (2018). How to detect insight moments in problem solving experiments // *Frontiers in Psychology*. Vol. 9. Art. 282. DOI: 10.3389/fpsyg.2018.00282
- Moroshkina N. V., Savina A. I., Ammalainen A. V., Gershkovich V. A., Zverev I. V., Lvova O. V. (2022). How difficult was it? Metacognitive judgments about problems and their solutions after the Aha moment // *Frontiers in Psychology*. Vol. 13. Art. 911904. DOI: 10.3389/fpsyg.2022.911904
- Salvi C., Bricolo E., Franconeri S. L., Kounios J., Beeman M. (2015). Sudden insight is associated with shutting out visual inputs // *Psychonomic Bulletin & Review*. Vol. 22(6). P. 1814–1819. DOI: 10.3758/s13423-015-0845-0
- Salvi C., Bricolo E., Kounios J., Bowden E. M., Beeman M. (2016). Insight solutions are correct more often than analytic solutions // *Thinking & Reasoning*. Vol. 22(4). P. 443–460. DOI: 10.1080/13546783.2016.1141798
- Salvi C., Simoncini C., Grafman J., Beeman M. (2020). Oculometric signature of switch into awareness? Pupil size predicts sudden insight whereas microsaccades predict problem-solving via analysis // *NeuroImage*. Vol. 217. Art. 116933. DOI: 10.1016/j.neuroimage.2020.116933
- Schooler J. W., Ohlsson S., Brooks K. (1993). Thoughts beyond words: When language overshadows insight // *Journal of Experimental Psychology: General*. Vol. 122(2). P. 166–183. DOI: 10.1037/0096-3445.122.2.166
- Shen W., Tong Y., Yuan Y., Zhan H., Liu C., Luo J., Cai H. (2018). Feeling the insight: Uncovering somatic markers of the “aha” experience // *Applied Psychophysiology and Biofeedback*. Vol. 43(1). P. 13–21. DOI: 10.1007/s10484-017-9381-1
- Shumilov T. V., Chistopolskaya A. V., Vladimirov I. Y. (2023). The Path to Insight: Developing Methods to Increase the Effectiveness of Insightful Solution Detection // *Psychology. Journal of Higher School of Economics*. Vol. 20(3). P. 473–489. DOI: 10.17323/1813-8918-2023-3-473-489
- Spiridonov V., Loginov N., Ardislamov V. (2021). Dissociation between the subjective experience of insight and performance in the CRA paradigm // *Journal of Cognitive Psychology*. Vol. 33(6–7). P. 685–699. DOI: 10.1080/20445911.2021.1900198
- Storme M., Myszkowski N., Celik P., Lubart T. (2014). Learning to judge creativity: The underlying mechanisms in creativity training for non-expert judges // *Learning and Individual Differences*. Vol. 32. P. 19–25. DOI: 10.1016/j.lindif.2014.03.002
- Tsai K. C. (2016). Quasi-experts’ assessments of creative products: An evaluation using a sample of design students // *British Journal of Education, Society & Behavioural Science*. Vol. 12(3). P. 1–7. DOI: 10.9734/BJESBS/2016/21057
- Valueva E. A., Lapteva N. M. (2023). “Aha!” and “Oh Yes!”: How emotions affect insight experience // *Psychology. Journal of Higher School of Economics*. Vol. 20(3). P. 428–444. DOI: 10.17323/1813-8918-2023-3-428-444